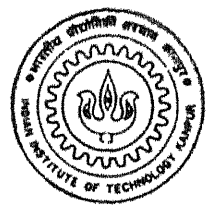
CLASSIFICATION OF PARTIAL DISCHARGE PATTERNS USING TEXTURE ANALYSIS ALGORITHMS

by

MOHAMED KAMAL ABD EL-RAHMAN



DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

JUNE, 1999

CLASSIFICATION OF PARTIAL DISCHARGE PATTERNS USING TEXTURE ANALYSIS ALGORITHMS

A Thesis Submitted
in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

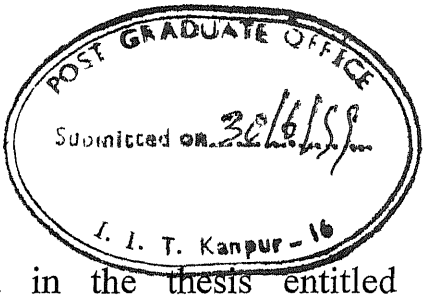
by

MOHAMED KAMAL ABD EL-RAHMAN

to the

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

JUNE, 1999



CERTIFICATE

It is certified that the work contained in the thesis entitled "**Classification of Partial Discharge Patterns Using Texture Analysis Algorithms**", by **Mohamed Kamal Abd El-Rahman**, has been carried out under our supervision. This work has not been submitted elsewhere for a degree.

R. Arora

Professor
Department of Electrical Engineering
Indian Institute of Technology
Kanpur, India

S.C. Srivastava

Professor
Department of Electrical engineering
Indian Institute of Technology
Kanpur, India

June, 1999

SYNOPSIS

Name of student	: Mohamed Kamal Abd el-Rahman
Roll no	: 9510462
Degree for which submitted	: Doctor of Philosophy
Department	: Electrical Engineering
Thesis title	: Classification of Partial Discharge Patterns Using Texture Analysis Algorithms
Name of supervisors	: Prof. Ravindra Arora & Prof. S C Srivastava
Month & Year of submission	: June, 1999

Partial discharge is an electric discharge that does not completely bridge the insulation between the electrodes. It gives rise to electric pulses having magnitude (q) and phase position (ϕ) with respect to the applied voltage waveform. It has been recognized that the breakdown of insulation of an electric equipment is often due to the occurrence of partial discharge within or on the surface of the insulation. Therefore, if partial discharge is found to occur in any insulation system, it is important to identify its source. The most important step in the classification process is to get the exact finger prints which could represent the different partial discharge sources successfully. During the past decade, partial discharge finger prints have been commonly formed by phase resolved pulse height analysis using the (q-ϕ-n) distributions, where n is the repetition rate of partial discharge

pulses Conventionally, classification of partial discharge sources has been done with the help of the shape of these distributions. The investigations in this thesis have been motivated by the fact that these distributions have the following drawbacks·

(a) They suffer from the averaging effect

(b) These distributions do not take the memory propagation between the partial discharge pulses into consideration.

Texture analysis algorithms have been successfully applied in the field of pattern recognition especially in the image processing applications, where the texture features contain information about the spatial distribution of spectral variation Generally, an image is divided horizontally and vertically into number of pixels. Each pixel has different gray level Texture analysis is used to study the gray level variation of an image in different directions (horizontal, vertical, left diagonal and right diagonal). To apply these algorithms to partial discharge source classification, if measurement is made through (say M) power frequency cycles and each cycle is divided into number of windows (say N), an image of $M \times N$ pixels is obtained In this study, the gray level has been replaced by the magnitude of partial discharge pulses so that the texture features can be used to give an overview of every single partial discharge pulse of the whole measurements. Investigations in this thesis have been conducted in the horizontal and vertical directions only in order to study the relationship between each pulse and the adjacent pulses in the same cycle as well as the relationship between a pulse and the other pulses in the same phase angle in different cycles.

This thesis has made an attempt to understand some of the issues in obtaining reliable features from texture analysis algorithms and evaluate their ability to classify different partial discharge sources The main objectives behind the research work carried out in this thesis have been to·

(a) compare the discriminating power of different texture analysis algorithms for the partial discharge source classification, including a minimum distance classifier, transformed divergence analysis and artificial neural network based methods.

(b) compare the discriminating power of the texture analysis algorithms with the discriminating power of the conventional (q-ϕ-n) distributions method for partial discharge source classification

(c) determine the relative classification accuracy of each feature individually.

(d) determine the minimum number of features used at a time to achieve a good classification accuracy.

(e) investigate the effect of changing number of cycles on the classification accuracy.

(f) determine the best feature selection technique in view of minimizing the number of features required to classify different partial discharge sources accurately.

(g) determine which direction (horizontal or vertical) has more discriminating power to minimize the computational time.

The work reported in this thesis has been organized into nine chapters:

Chapter 1 introduces the partial discharge phenomena, reviews the conventional methods for partial discharge classification, and briefly outlines the work carried out in this thesis.

Chapter 2 presents theoretical background of four different texture analysis algorithms. These algorithms include the spatial gray level dependence method (SGLDM), the gray level difference histogram method (GLDHM), the gray level run length method (GLRLM), and the power spectral method (PSM), respectively.

Chapter 3 describes the experimental set up used for partial discharge measurements and the samples used to simulate the partial discharge sources. The partial discharge sources created in the laboratory include glow corona, streamer corona, surface discharge, internal discharge, single protrusion and multi protrusions.

Chapter 4 describe the application of the minimum distance classifier for partial discharge classification using all the features based on the texture analysis algorithms. The classification accuracy of each algorithm has been investigated and the classification accuracy of each feature individually as well as for the combination of two features at a time have also been studied.

Chapter 5 presents the application of a direct feature selection technique known as the 'transformed divergence analysis' to measure the distance between the classes in the

feature space. The optimum features have been selected according to the maximum separation between the different partial discharge sources.

Chapter 6 investigates the use of a back propagation artificial neural network (ANN) for partial discharge classification. A three layers feed forward ANN has been developed for this purpose.

In chapter 7, the GLDHM features have been used to distinguish two partial discharge sources generated from a high voltage cable as a practical example. A belted cable has been used to generate internal discharge between the insulating paper and also the surface discharge at the terminals of the cable.

In chapter 8, an indirect method known as 'principal component transformation' has been used to reduce the number of features by mapping the original set of features into a new set of features with less number in which the separation between the classes (partial discharge sources) are maximum. Classification accuracy of the main principal components has been determined for different partial discharge sources.

Chapter 9 presents summary of main conclusions of the thesis and also includes few recommendation for further work.

The work carried out in this thesis reveals the following:

1. Texture analysis algorithms can be effectively used to generate different features which are capable of distinguishing between different partial discharge sources.
2. Amongst the four texture analysis algorithms, GLDHM has a classification accuracy at par with the other algorithms and is computationally faster. Therefore, it is recommended to be used for partial discharge source classification.
3. With the use of minimum distance classifier and also the transformed divergence analysis for each of the algorithms, two features used at a time are found sufficient to achieve a considerable classification accuracy. However, the best combinations depend upon the number of cycles used to construct the patterns and also on the partial discharge sources used for classification.
4. The classification accuracy achieved with the ANN is less than that achieved with the minimum distance classifier.
5. Using the principal component transformation, two principal components are found to be sufficient to achieve a desired classification accuracy which is found to be independent of the number of cycles used to construct the partial discharge patterns and also the partial discharge sources used for classification. Hence, the principal component transformation is recommended as a technique to reduce the number of features for the partial discharge source classification.

ACKNOWLEDGMENT

I express my deep sense of gratitude to **Dr. RAVINDRA ARORA**, Professor, Department of Electrical Engineering, Indian Institute of Technology, Kanpur and **Dr. S. C. SRIVASTAVA**, Professor, Department of Electrical Engineering, Indian Institute of Technology, Kanpur for their expert guidance and unreserved co-operation which enable me to bring out this thesis in the present form successfully.

My sincere thanks to **Dr. ONKAR DIKSHIT**, Department of Civil Engineering, Indian Institute of Technology, Kanpur, for his cordial help in different ways throughout the work.

My sincere thanks to all persons who directly or indirectly contributed in progress of this work.

At the end I wish to thank all **MY FAMILY MEMBERS** for their loving attention.

M. K. Abd El-Rahman

To

My Wife

CONTENT

CHAPTER 1

Introduction

Chapter 1

Introduction

1.1 General Introduction

Electric discharges which do not bridge the electrodes are called partial discharge [Kreuger, 1989]. In other words, it is a local breakdown in a dielectric when it loses its insulating property locally and not globally. It may originate in the dielectric directly at one of the electrodes or occur in a cavity. The partial discharge phenomenon results in very short duration electric pulses. Although the magnitude of such pulses is small, they can cause progressive deterioration and ultimate failure to the insulation. Therefore, it is essential to detect their presence in order to ensure reliable operation of high voltage equipment. Furthermore, a meaningful interpretation of partial discharge measurement is required in order to identify its source. Several methods are available for partial discharge detection and classification. This chapter introduces some of the existing approaches for partial discharge detection and classification along with review of some of the literature on the topic and set the motivation behind the present work.

1.2 Types of Partial Discharge

Partial discharges are categorised into the following three groups[Kreuger, 1989]:

(1) *Corona or gas discharge*: Partial discharge at a free electrode in gaseous dielectric is

known as corona. Depending upon the electrode shape and the gap distance, these can be of three types namely glow, streamer and leader corona.

(2) *Surface discharge,* also knowing as tracking, which occurs on the interface of solid or liquid dielectric materials with gaseous dielectrics. Surface discharge may occur if there exists an electric stress component in the direction of the dielectric surface like in bushings, insulators and ends of cables.

(3) *Internal discharge*: Internal discharges in a dielectric occur due to foreign particle inclusions during manufacturing or operation as well as in cavities. Cavities could be filled with gas or oil. The dielectric strength of the material in the cavities breaks down at a stress which is much lower as compared to the breakdown strength of the surrounding dielectric. The dielectric strength of the gas filled cavities depends on :

(i) The kind of gas in the cavity

(ii) The gas pressure

(iii) The shape and direction of the cavity.

Oil filled cavities occur between layers of oil impregnated paper insulation such as in transformer winding and in cables. The breakdown strength of oil strongly depends upon contamination and the amount of the dissolved gases. If the oil breaks down, gas bubbles are produced and gas discharge may occur.

1.3 Recurrence of Discharge

When the voltage across a cavity reaches the breakdown value, the cavity may break down. The voltage breakdown takes place in less than 10 μsec [Kreuger, 1989]. This is extremely a short time compared to the duration of a 50 Hz sinewave and hence the voltage drop may be regarded as a step function. The breakdown of the cavity is determined by the superposition of the main electric field and the field of the surface charge at the cavity walls left behind after the last discharge. The voltage across the cavity starts again increasing until it reaches the breakdown voltage, when a new discharge occurs. Thus several discharges may take place during the rising part of the applied voltage. Similarly, on the decreasing part of the applied voltage the cavity discharges occur as the voltage across the cavity reaches the negative value of the breakdown value. In this way, groups of discharges originate from a cavity and gives rise to positive and negative current pulses on raising and decreasing the voltage, respectively, as shown in Fig. 1.1.

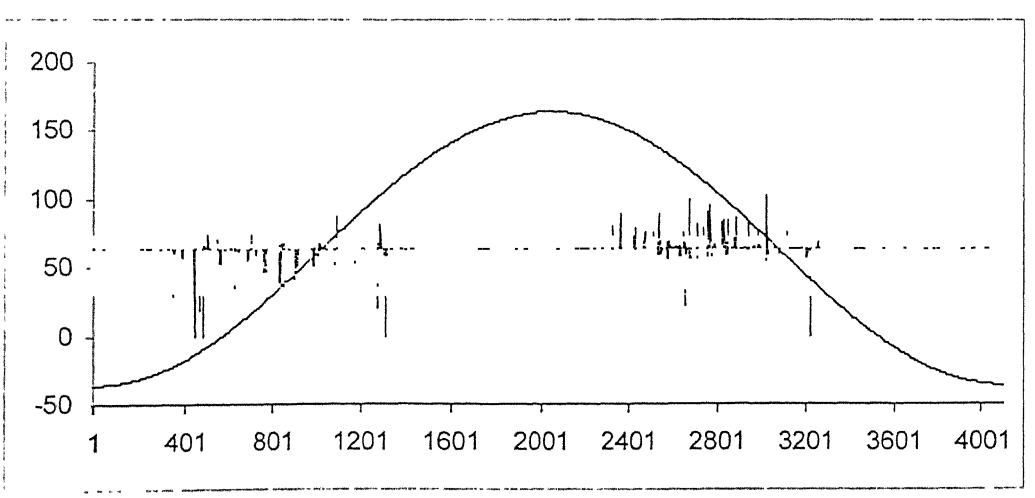


Fig. 1.1: Partial discharge pulses with respect to the applied voltage

1.4 Detection of Discharge

The detection of discharges is based on energy exchanges which take place during the discharge. Electrical discharge is transitory disturbance which radiates electromagnetic, acoustic and thermal energy from the discharge site. Therefore, these exchanges are manifested as electrical impulse current, dielectric loss, chemical transformation, heat, gas pressure, sound and light [Kuffel *et al*, 1984]. Generally, there are two types of partial discharge detection methods; non electrical and electrical methods.

1.4.1 Non-electrical detection methods

Both insulating materials and lubricating oils are complex organic materials which, when degraded by heat or electric action, produce a very large number of chemical products in the gas, liquid and solid state [Tavner *et al*, 1987]. Electrical discharge activity within or adjacent to the insulating system also release chemical degradation products. It breaks oxygen to give ozone. Furthermore, continuos partial discharge activity gradually carbonise the insulating materials to produce, on a smaller scale, the degradation products which, for example, are caused by local overheating.

Sound is a longitudinal mechanical wave motion in an elastic medium and is classified according to its frequency as [Tuma, 1976]:

(i) *audible* by human ear (frequency between 20-20000 Hz)

(ii) *infrasonic* (below the response of the human ear)

(iii) *ultrasonic* (above the response of human ear)

Sound detection could be audible or ultrasonic. Usually a narrow band of about 30 to 50 kHz is chosen (just above the audible spectrum). Several ultrasonic detection systems are available commercially. The signals are converted to audible sound and their magnitude can be read from a decibel meter. Ultrasonic techniques are useful for the detection and location of partial discharge sources, especially for oil filled transformers. Investigation of ultrasonic frequency spectrum from 0.15 to 1.75 MHz has revealed that the ultrasonic spectrum signature of some partial discharge sources, like voids and spark gaps, are different and thus can aid in partial discharge identification [Harrold, 1975].

Light detection can only be used for surface discharge and corona. The sensitivity of light detection can be improved appreciably by using a photomultiplier [Kreuger *et al*, 1988]. However, it shares the disadvantage with other non electrical methods, that the discharge magnitude can not be measured. The non electrical detection methods are not used commonly because they are less sensitive than the electrical ones.

1.4.2 Electrical detection methods

The primary characteristics which partial discharge detectors have in common and, therefore, could provide a basis for evaluation of the detector are the number of inputs employed, the bandwidth of the detector and the method of display processing [Steiner, 1991]. The three basic types of displays are:

(i) The meter display

(ii) The direct display

(iii) The computer driven display.

The meter methods are implemented using either a digital panel meter or an analog meter

and display a number related to some parameter of the partial discharge process. Also, more than one meter can be used as an auxiliary display device so that several parameters can be monitored simultaneously. The main disadvantage of meter display is that it is not able to distinguish between a single discharge of highest magnitude and the sum of multiple discharges of smaller magnitudes. The long time-constant of the voltmeter, for example, causes it to behave somewhat like an integrator providing only measure of an average value[Steiner, 1991]. Direct display is the best choice. The discharge impulses usually are displayed on a time base of the same frequency as the applied voltage. Recurrent discharge in successive cycles cover each other and a stationary picture is obtained. From the pattern of a discharge on the screen of the oscilloscope, the source of that discharge can be determined. Computer driven display requires that the detected partial discharge pulses be digitised either into a set of values or waveforms. The most common partial discharge electrical detectors are:

1.4.2.1 Schering bridge

The occurrence of partial discharge in electric equipment was already recognised by Peterson at the beginning of this century in 1912. Its presence has been normally detected either visually or audibly at sufficient high intensity. Following the development of Schering bridge in the 1920's, it became possible to record or detect discharge in terms of its influence upon the dissipation factor (*tan* δ) value. The power loss P in a dielectric can be expressed by

$$P = \omega C V^2 \tan\delta$$

where ω is the frequency, C the specimen capacitance, V the applied voltage and *tan* δ is

the dissipation factor. It was noted that a plot of *log(P)* versus *log(V)* is straight line having a slope equal to two so long as the tan δ value remains approximately constant with the applied voltage [Bartnikas, 1987]. Following PD inception with increasing voltage, the increase in *tan δ* with voltage becomes perceptible, and consequently the slope of the line exceeds the value of 2. Hence, after partial discharge inception, the plot will consists of two straight lines whose intersection will determine the partial discharge inception voltage. Schering bridge is a narrow band detector and normally implemented as a two input detector. Since the bandwidth is small, an integrated response is provided. Therefore, normal background partial discharge masks the presence of larger pulses. It is not accurate due to the following reasons:

1- Not every increase in dielectric losses coincides with discharges.

1- Start of increase in *tan δ* is often difficult to determine.

2- The amount of losses due to partial discharge is quite negligible as compared to the other power losses in the dielectric.

1.4.2.2 Radio interference voltage

The Radio Interference Voltage (RIV) is one of the oldest partial discharge measurement method. This measurement technique finds its origin in Electromagnetic Interference Measurements (EMI) and was not designed but adopted for the partial discharge measurement. The RIV method is still common in the transformer industry and also used for measuring partial discharge in bushing and insulators [Steiner, 1991]. RIV is a narrow band single input detector. Therefore, its long time- constant causes it to behave somewhat

like an integrator providing only a measurement for the average value of the partial discharge magnitude. The main disadvantage of RIV is that no general relationship exists between the measured voltage and the average partial discharge level in the equipment.

1.4.2.3 Single channel analyser

The partial discharge detecting techniques described earlier can be classified as (go-no go) methods in the sense that they only indicate the presence or absence of discharge above certain levels in pico Coulomb (pC) or micro Volt (μV) in the insulating system undergoing test. Additional quantitative information on the partial discharge process concerning the partial discharge pulse pattern density was obtained when measurements were made for the discharge rate. This could be done by using electronic counters to count partial discharge pulse amplitudes above certain pre-set voltage or by the additional use of discriminator to count pulses within only certain level. This method is commonly known as single channel analyser [Bartnikas *et al*, 1969]. With the help of the single channel analyser, measurements are carried out for the discharge rate as a function of the partial discharge magnitude. This could be done by either varying the levels of the upper and lower discriminators or by making use of a limited number of single channel units with different but fixed level settings. By changing the level of the upper and lower discriminator, the single channel analyser proved to be susceptible to error due to time dependent changes in the partial discharge patterns itself, whereas the use of limited number of single channel analysers was characterised by extremely poor pulse height resolution characteristics.

1.4.2.4 Multi-channel analyser

With the availability of multi-channel pulse height analyser, the above difficulties could be resolved. A multi-channel pulse height analyser provides essentially a distribution of the frequency of occurrence of the partial discharge pulse heights of a train of pulses associated with a given partial discharge pattern [Bartnikas, 1973a; Bartnikas, 1973b]. Commercially available pulse-height analysers may have a typical options of 128, 256, 512, 1024, 2048, 4096, 8192 channels coupled with the required memory and capability of digitising the input pulse data. For example, in 1024 channel pulse height analyser, the incoming partial discharge pulses of various amplitudes are sorted by an analog to digital converter into one of 1024 possible heights. Thus, multi-channel analyser provides a statistical distribution characteristics in which individual channels 1, 2, 3,..,1024 correspond to particular pulse charges. The total number of partial discharge pulses counted in each channel equals the number of discrete pulses whose peak amplitude corresponds to the particular channel. The count time interval over which each partial discharge pulse train is analysed can be made sufficiently long to obtain a truly statistical amplitude distribution of the pulse train. Each channel provides a point reading and a curve drawn through the points represents a plot of the number of partial discharge pulses versus the apparent charge. *RC* pulse shaping circuit can be employed to provide the multi-channel pulse height analyser with smooth unidirectional pulses. The horizontal axis of the multi-channel analyser has to be calibrated in μV or pC i.e. each channel corresponds to a certain voltage threshold or level. Prior to each measurement the over all amplification of the corona detection circuit must be suitably adjusted to fix the upper and lower amplitude spectrum limits between which a particular partial discharge pulse train is to be analysed. By using a multi-channel pulse

height analyser one could get a family of partial discharge pulse height distributions as a function of applied voltage or time.

However, it should be noted that the multi-channel analyser performs analysis of unipolar pulses only. Therefore, pulse amplitude distributions for positive and negative discharges can only be obtained by two distinct measurements carried out at different times. Moreover, it does not make efficient use of the available information since it filters the data received and does not provide a complete permanent record of all partial discharge events during the time of observation. On the other hand, multi-channel analyser is important in DC testing because these pulses are difficult to observe if there is no means available to store them. DC partial discharges are difficult to observe on a direct display because the time separation of pulses are random and quite often these time spans are large.

1.4.2.5 Pulse detection circuits

Broadband pulse detection is by far the most common method for measuring partial discharge. The typical pulse detection circuit is implemented using a high voltage coupling capacitor and a measuring impedance. The coupling capacitor must be partial discharge free up to the maximum test voltage of the system. The measuring impedance may be connected to the sample in two ways:

(i) In series with the test sample.

(ii) In series with the coupling capacitor.

Both the methods are electrically equal, where the same partial discharge pulses pass across

the measuring impedance. But, in practice, it is often placed in series with the coupling capacitor because of the following reasons:

(i) The large charging current of the test sample does not pass through the measuring impedance.

(ii) To protect the measuring impedance and measuring circuit from over voltage in case of the test sample failure.

(iii) It is particularly useful when it is not practical to break the ground connection to the test sample.

The measuring impedance used commonly are:

(i) A resistor shunted by a parasitic capacitor (*RC* impedance).

(ii) An oscillatory circuit (*RLC* impedance).

The higher sensitivity of *RLC* type pulse detection method was recognised as early as in 1933. It is used essentially as a resonant type of corona pulse detection circuit which was set into oscillation at its natural frequency by a discharge transfer within the test sample. The capacitance of the measuring circuit is small. It just represents a stray capacitance. Also the inductance of the measuring circuit has a low impedance at power frequency so that the AC excitation voltage is eliminated across the coupling capacitor. The *RLC* impedance acts as a band-pass filter and transforms the very short partial discharge pulses, superimposed on the power frequency component, into decaying cosine transient pulses. A

high resonance frequency of the measuring impedance is required to ensure high pulse resolution time and therefore integrated pulse having short duration, but at the same time a small value of the upper cut off frequency is necessary to perform a good current pulse integration. The lower cut off frequency is suggested to be greater than 10 kHz. The upper cut-off frequency ranges from 200 to 400 kHz to limit high noise interference with the detected signal [Capponi *et al*, 1992]. The high frequency noise is picked up from commercial AM radio transmissions which are in the frequency band of 540 to 1600 kHz. The factor that influences the choice of the lower cut-off frequency is the need to eliminate low frequency interference such as the harmonic distortion in the power source and switching noise.

The direct display method is most common method for displaying broadband partial discharge pulses. Broadband partial discharge detection systems are widely used as a result of several advantages compared to narrow band techniques. The most important advantage is that the individual partial discharge pulses are measured with this method rather than an average value. Another important advantage is the greater sensitivity which comes with a wider bandwidth. With pulse detection systems, the operator can directly observe the pulse shape along with the pulse polarity. This is not possible with narrow band systems. The operator can also observe the distribution of pulses in phase, which is often useful in identifying the nature of the discharge source. These types of measurements are readily adapted by the computer controlled systems.

For partial discharge measurements, following two fundamental methods could be used:

(i) Balanced detection bridge method

(ii) Direct measurement method.

In the first case, partial discharge signals are obtained as difference between the voltages across two resistive impedances put in two distinct branches, one containing the specimen under test and the other a discharge free coupling capacitor having the same frequency dependence of dielectric loss as the specimen. In this case, the displacement current and its harmonics are suppressed by common rejection mode of the bridge circuit. However, it should be pointed out that this method is seldom used because of some drawbacks. Actually this method requires the specimen to be kept above the ground and, in order to get the full advantage of the high value of common rejection mode ratio, a coupling capacitor must be available, matching the specimen as far as the capacitance and dissipation factor are concerned. These conditions can not always be achieved in partial discharge testing routines.

In the direct measurement method, current pulses in the leads of the sample are measured directly. One of the main difficulties in performing partial discharge tests is that the discharges which occur in other parts of the test circuit, such as in the high voltage source, high voltage leads, coupling capacitor etc interface the measurements. The impulses caused by these discharges are difficult to be distinguished from those discharges in the test object, so that they disturb the observation of the wanted discharges. The effect of these external discharges can partially be suppressed by using filters or by using partial discharge free high voltage transformer and coupling capacitor.

1.5 Computerised Measurement and Automated Recognition of Partial Discharge

The data obtained using various multi channel pulse analyser techniques can be stored and then analysed by means of a computer. The current trend in this regard is to substitute

much of multi-channel analyser hardware with computer oriented software. One of the advantages of this method is the possibility of later improvement of the evaluations, since the primary data are on permanent record in storage media. This is in contrast to the earlier techniques when a new evaluation required application of high voltage again to the test object which might have changed with time.

For many years, recognition of partial discharge pattern was performed by visual inspection, i.e. by observation of partial discharge on an oscilloscope screen. The interpretation of these patterns appearing on the ellipse depends upon the knowledge and experience of the expert. It became soon apparent to the worker in this field that the measurement of total or cumulative pulse count was not capable of providing a complete overview of the process occurring during the partial discharge. A measure of the discrete pulse height with their corresponding discharge rate was very much desirable. Therefore, in addition to the usual measurement of the partial discharge inception and extinction voltages as well as the maximum partial discharge pulse amplitude, it is useful to record other partial discharge parameters, such as (for example) the maximum pulse magnitude, the pulse repetition rate, the pulse separation time and the pulse relationship with respect to the applied sinusoidal voltage. The aim of today's computer diagnosis system is to replace the human expert. However, this requires a better diagnosis reliability. For visual display the dimension is limited to only three parameters. Partial discharge events have a magnitude and a phase position. They also depend upon time and the test voltage. Generally a personal computer is able to handle as many quantities as necessary. Also, once the raw partial discharge data from the interfacing circuit has been transferred into the computer memory, the information can be analysed immediately and displayed as required. Therefore, it is

useful to develop more complex distinguishing features during the automated recognition of partial discharge phenomena.

The development of computer based partial discharge measurement technique was started in UK [Austin *et al,*1976]. In their work, for the first time, the discharge information associated with individual partial discharge pulse was recorded digitally using a mini computer. Such information included the apparent discharge magnitude, its polarity, time of occurrence and the applied voltage. The use of computer in partial discharge measurements started the automation in partial discharge recognition [Gulski *et al* 1990, 1992 & 1995; Gulski 1993, 1995a &1995b; Krivda 1995a]. As each defect has its own particular effect on the degradation of insulation, it is important to know the correlation between the discharge pattern and the kind of defect. Therefore, progress in the recognition of partial discharge pattern and their correlation with the kind of defect is becoming increasingly important in the field of quality control of insulating systems. One of the most important advantages of a computer-aided measuring system is the ability to process a large amount of information and to transform it into an understandable output form.

Many computer-aided systems have been developed for the measurement and understanding of partial discharge phenomena. In practice, the trend is in the direction of improving the recognition of discharge sources and evaluation of the measuring results which assist in determining the quality and the condition of the insulation system. In this regard, the phase resolved pulse height analyser is the most common approach [Okamoto *et al* 1982, 1985]. In this approach, pulse height analysis, which was common with multi-channel analyser, is performed as a function of the excitation phase voltage. Phase resolved pulse height analysis has several advantages in addition to the usual benefits of the

broadband methods. The first advantage is that the measurements are stored in memory. Since other methods do not have memory, measurements are lost. After the data collection, further analysis can be performed such as signature analysis and various statistical operations. However, the information regarding the actual time of occurrence of each individual partial discharge pulse relative to the start of the test is lost after this conversion.

1.6 Recognition Procedures

In general, recognition procedures are shown in Fig.1. 2:

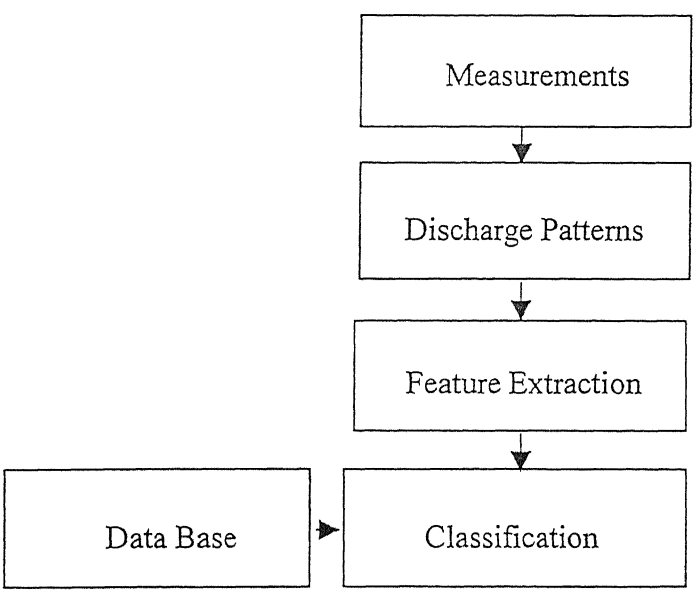


Fig. 1.2 Pattern recognition procedure

Various modules of the recognition procedure, shown in the above figure, are described below for the partial discharge source classification.

1.6.1 Measurements

Partial discharge measured quantities can be divided into two main groups:

1- Basic quantities, which are observed during single voltage cycle .

2- Deduced quantities, which are derived from basic quantities in the first group observed throughout several voltage cycles.

1.6.1.1 Basic quantities

It is known that by using conventional detection methods (bandwidth 400 kHz) the electrical activity of partial discharge is represented by the following independent quantities [Kreuger *et al*, 1993; James *et al* 1995]:

- Discharge magnitude q_i
- Instantaneous applied voltage v_i
- The position of the discharge related to the phase angle of the test voltage ϕ_i.
- Time of occurrence relative to the beginning of test t_i.
- Discharge polarity p_i (positive or negative).
- Discharge energy w_i.

1.6.1.2 Deduced quantities

For obtaining the deduced quantities, the basic quantities have to be observed for a long time span. These quantities can be analysed as a function of time as well as the phase angle of the applied voltage. The quantities as function of time describe the change of the basic quantities with respect to time. It is known that variations in partial discharge occur, both

in magnitude and also in temporal behaviour of the discharge. This variation is due to statistical variations in the discharge phenomena itself and change in the discharge site. Therefore, to get information on the condition of the dielectric, the time behaviour in both the positive and negative half cycles of inception voltage $V_{ins}(t)$ and the number of discharge $N_q(t)$ were processed [Gulski *et al*, 1992]. The quantities as a function of phase angle represent the recurrence of partial discharge related to their phase angle. Therefore the voltage cycle was divided into phase windows representing the phase angle axis (0-360°). If the observation takes place over several voltage cycles, the following four quantities can be determined in each phase window [Gulski *et al*, 1992; Kreuger *et al*, 1993]:

- Sum of the discharge magnitudes
- Number of discharges
- Average value of discharges
- Maximum value of discharge

These quantities observed throughout the whole angle axis, result in distributions of recurrence as function of phase angle.

1.6.2 Patterns for partial discharge recognition

Partial discharge measurements can be performed in many ways. The measurements give rise to patterns. By measuring charge displacement in the leads, in the form of current pulses, partial discharge patterns can be observed in the form of various discharge distributions. These distributions can be three dimensional Hn(q,ϕ) and the other two dimensional distributions derived from it are as following:

- $Hn(q,\phi)$: Number of pulses as function of magnitude and phase angle.
- $Hqmax(\phi)$: The maximum pulse height distribution as function of phase angle.
- $Hqm(\phi)$: The mean pulse height distribution as function of phase angle.
- $Hn(\phi)$: The pulse count distribution as function of phase angle.
- Hn(q) : The number of discharge as a function of discharge magnitude.
- Hn(p) : The number of discharges as a function of discharge energy.

These distributions have proved to be useful for the recognition of partial discharge sources. It has also been observed that these distributions can significantly change during ageing of insulation [Krivda *et al*, 1994; Gulski *et al*, 1995].

1.6.3 Feature extraction for partial discharge recognition

The aim of feature extraction is to reduce the dimensionality of the original partial discharge patterns by calculating certain features from the patterns. The number of features should be as low as possible. Lower the number of features, faster is the speed of classification. Most of the features are based on the three dimensional distribution $Hn(q,\phi)$. The number of pulses in each window of the $Hn(q,\phi)$ distribution, measured for certain partial discharge source, can be used to generate number of features equal to the number of windows which can be used in turn to identify its source. However, by using high resolution distribution, i.e. large number of phase windows and magnitude windows, a large number of features, have to be used for recognition. Even by using a low resolution, the number of features required is high. To reduce the number of features to a reasonable number, the following methods have been used in the literature.

1.6.3.1 Statistical features

All the data which have common feature that they are affected by chance should be analysed by statistical operators. Situations may get influenced by the presence of such effects which one can not predict because they result from factors that can not be controlled or often enumerated. Partial discharge measurements are example of the above kind of data. Therefore, different statistical parameters such as skewness, kurtosis asymmetry etc. have been used to characterise the partial discharge distributions.

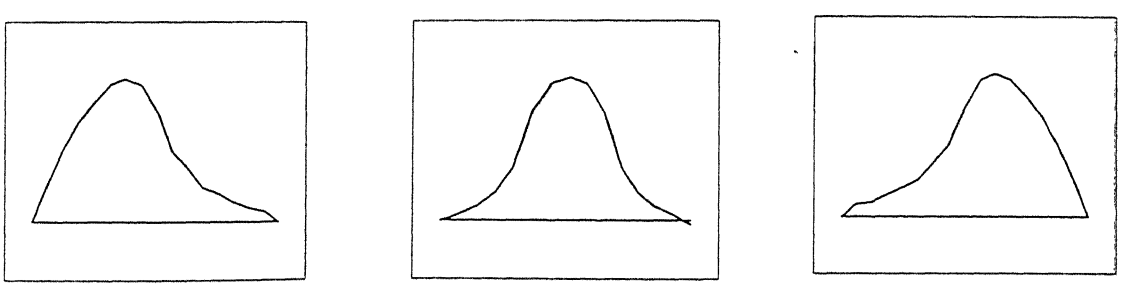
(i) Skewness (Sk)

It is defined as:

$$Sk = \frac{\sum (x_i - \mu)^3 p_i}{\sigma^3} \qquad (1.1)$$

where x_i is the recorded value, p_i is the probability of frequency of appearance for that value x_i in phase window i, μ is the mean value and σ is the standard deviation. It represents the asymmetry of the distribution as shown in Fig.1.3

- If the distribution is symmetric, $Sk = 0$
- If it is asymmetric to the left, $Sk > 0$
- If it is asymmetric to the right, $Sk < 0$



$Sk < 0$ $Sk = 0$ $Sk > 0$

Fig. 1.3: Distributions with different skewness

(ii) Kurtosis

It represents the sharpness of the distribution. It is defined as:

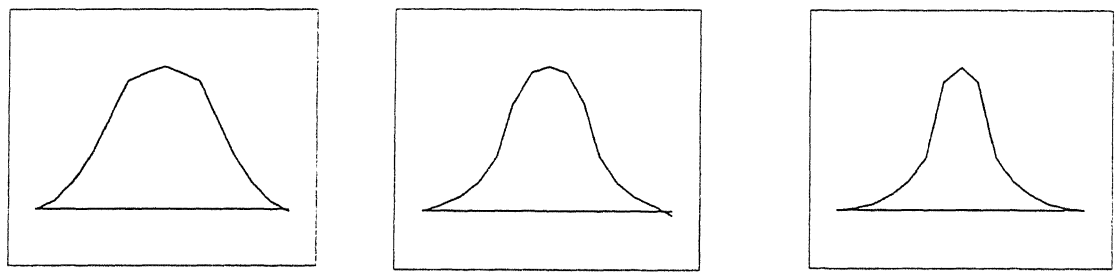
$$Ku = \frac{\sum (x_i - \mu)^4 p_i}{\sigma^4} \tag{1.2}$$

As shown in Fig. 1.4,

- If the distribution has the same sharpness as the normal distribution, $Ku = 0$

- If it is sharper than the normal distribution, $Ku > 0$

- If it is flatter than the normal distribution, $Ku < 0$



Ku < 0 *Ku=0* *Ku >0*

Fig. 1.4: Distributions with different kurtosis

(iii) Cross-correlation factor (Cc)

The cross correlation factor indicates the difference in shape of partial discharge pattern in the positive and negative half cycles. If the shapes are the same (but not necessarily equal) *Cc*=1. If they differ completely, *Cc*=0. It is defined as,

$$Cc = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sqrt{\left[\sum x_i^2 - (\sum x_i)^2 / n\right]\left[\sum y_i^2 - (\sum y_i)^2 / n\right]}} \tag{1.3}$$

where x_i is the mean discharge magnitude in window i of positive half cycle and y_i the mean discharge magnitude in the corresponding window in the negative half cycle; n is the number of phase windows per half cycle.

(iv) Discharge factor (Q)

The discharge factor describes the difference in the mean discharge levels in the positive and negative distributions. It is defined as,

$$Q = \frac{Q_s^-}{N^-} \Big/ \frac{Q_s^+}{N^+} \tag{1.4}$$

where,

Q_s^+ and Q_s^- are the sum of discharge of the mean pulse height distribution in the positive and negative half cycles, respectively.

N^+ and N^- are the number of discharge of the mean pulse height distribution in the positive and negative half cycles, respectively.

Q=1 denotes equal discharge and Q=0 denotes the large difference

(v) Modified cross correlation factor (MCc)

The modified cross correlation factor is defined as the product of the discharge factor and the cross correlation factor.

$$MCc = Q \,.\, Cc \tag{1.5}$$

(vi) Number of peaks

It is used to distinguish between the distributions with single peak and distributions with

several peaks.

To improve the discrimination of these features, it has been suggested that these should be calculated for the positive and negative half cycles, separately. For example, skewness, kurtosis and number of peaks can be calculated for each half cycle individually, while the cross correlation factor can be calculated to investigate the correlation between the distributions in the positive and negative half cycles. For example in this case, for the three distributions, Hn(φ), Hqm(φ) and Hqmax(φ) one gets 21 features.

1.6.3.2 Fractal geometry

Fractal provides a proper mathematical framework to study the irregular and complex shapes e.g. tree, hill, clouds, etc. as found in nature. The geometry of fractals has been used to extract features from the three dimensional partial discharge patterns acquired using phase resolved partial discharge analyser (Satish *et al*, 1995; Krivada, 1995a; Krivada *et al*, 1995b; Meijer *et al*, 1998). Following two fractal features have been used:

(i) Fractal dimension to characterise the surface roughness.

(ii) Lacunarity to describe the surface denseness.

The calculation of both these features depends upon the number of cubes of side l which is necessary to cover the fractal surface. To calculate this number, it is assumed that $p(m,l)$ is the probability that there are m points within a cube of size l which is centred about a point on the fractal surface. $p(m,l)$ is normalised , as below, for all l.

$$\sum_{m=1}^{N} p(m,l) = 1 \tag{1.6}$$

Where, N is the number of possible points within the cube. Let S be the total number of surface point. If one overlays the surface with cubes of side l, then the number of cubes with m points inside the cube is $(S/m)\, p(m,l)$. Therefore, the expected total number of cubes to cover the whole surface is:

$$N(l) = \sum_{m=1}^{N} \frac{S}{m} p(m,l) = S \sum \frac{1}{m} p(m,l) \tag{1.7}$$

To calculate the fractal dimension (D), $N(l)$ should be calculated for various values of l using a least square curve fitting on $[log(l), log(n(l))]$, where D is the slope of that line. Lacunarity $\Lambda(l)$ could be calculated by using the second order statistics of $p(m,l)$ as following:

$$M(l) = \sum_{m=1}^{N} m p(m,l) \tag{1.8}$$

$$M^2(l) = m^2 p(m,l) \tag{1.9}$$

$$\Lambda(l) = \frac{M^2(l) - [M(l)]^2}{[M(l)]^2} \tag{1.10}$$

1.6.4 Classification

The aim of classification is to assign a label to a partial discharge pattern of unknown origin from previously collected patterns with known labels such as treeing discharge, corona etc. There are number of methods available for classification. Some of these include:

- Artificial neural network based classifiers (ANN)

- Conventional classifiers

1.6.4.1 Artificial neural networks based classifiers

The classification potential of an ANN can be visualised by understanding how it classifies the finger prints. The back-propagation network (with one hidden layer and a sigmiod transfer function) separates data by hyperplanes (line in 2-d space, planes in 3-d space, etc.). The hyperplanes are generated by neurons in the hidden layer (one hyperplane per neuron). Weight connections between the input layer and the hidden layer determine the slope and shift of the hyperplanes. Weight connections between hidden layer and the output layer serve as logical functions which decide on which side of hyperplane a test fingerprint lies. Finger prints of two defects can be separated by the network with two neurons in the hidden layer. A testing finger print is then classified according to its position relative to hyperplane. Such a classification procedure can, however, cause problems. It can be seen that a fingerprint of unknown origin, not belonging to any of these defects, could be classified to a certain defect because the unknown pattern and that defect lie on the same side of the hyperplane. More neurons are therefore required in the hidden layer to separate the fingerprints of the two defects from the surrounding space. However, in more than two dimensions, the structure of data is unknown and it is difficult to estimate the number of neurons in the hidden layer. Also, when it is required to add new defects for recognition to a previously trained network, the network has to be completely retrained. The ANN learning process is, in general, time consuming.

Although the application of ANNs is reasonably established, its application in the area of partial discharge is relatively new. An important aspect in the application of ANNs is the

definition of the input information which is fed into the ANNs. The following paragraphs briefly review the ANNs and the type of the input patterns used which have been used for the partial discharge classification.

A three layer feed forward ANN with back propagation learning algorithm has been used for the automatic distinguishing between partial discharge generated in an XLPE cable and noise (Suzuki *et. al.* 1992). A sample of 5 m length of cable was used to generate different patterns with and without partial discharge. Partial discharge was generated by using a metallic needle electrode with a tip radius of 5μm which was inserted 1 *mm* into the insulation of the cable. The input patterns were related to the number of pulse counts (say *n)* in each pixels of the ϕ–*q-n* distribution. To reduce the input data, the ϕ–*q-n* distribution was divided into a 20 (phase angle) × 32 (discharge magnitude) pixels. In addition, pixels for negative pulses in the positive half cycle and positive pulses in the negative half cycle were ignored to reduce the insignificant pixels. Therefore, the input patterns of the ANN consisted of a series of 320 input values. Three types of input patterns; ϕ–*q-1*, ϕ–*q-n*, and ϕ–*q-n*` were used. In case of $\phi$–*q-1* pattern, the input value was taken to be equal to one if the number of pulses was greater than or equal to a given threshold value. In case of the $\phi$–*q-n*, the input value was equal to the number of pulse counts. In the case of $\phi$–*q-n*`, the input value was equal to the actual number of pulses in the corresponding pixel plus a correction factor calculated from the adjacent 8 pixels in order to smoothen the patterns. It was found that the ANN could easily discriminate partial discharge from noise by using ϕ–*q-1* patterns.

ANN using back propagation method has been applied for the discrimination of partial

discharge patterns before and after the tree initiation from the needle weak point [Hozumi *et al*, 1992]. A metallic needle was inserted in epoxy resin to initiate the electric tree. The input patterns were based on the ϕ–q-n and ϕ–q distributions. The ϕ–q-n was divided into 20 windows for the phase angle and 16 windows for the partial discharge magnitude. Therefore, a ϕ–q-n pattern was composed of 320 values. The ϕ–q indicates the mean partial discharge magnitude of the pulses in each phase window. The number of windows was 20, the width of the phase window being 18°. However, the network which was learnt from ϕ–q-n patterns showed better discrimination performance than that which was learnt from ϕ–q patterns. Similar technique with the same input patterns was used for seven different partial discharge source classification by using three dimensional patterns [Satish *et al*, 1994].

Another approach based on 15 statistical parameters like skewness, kurtosis, number of peaks and cross correlation had been used to define input pattern of the ANN [Gulski *et al*, 1993]. These parameters were used to quantify the profiles of two phase position distributions: *the mean pulse height distribution* (q-φ) which shows the average partial discharge magnitude in each phase window as a function of the phase angle φ and *the pulse count distribution* (n-φ) which shows the number of partial discharge in each phase window as a function of the phase angle φ. Back propagation network, Kohonen self organising map and learning vector quantization network were used for classification in this work. Two ways of applying ANN were studied: (i) In the first case, single ANN was used with number of outputs equal to the number of partial discharge sources studied. (ii) In the second case, for each partial discharge source, a separate ANN with one output was built. Thus in the second case M ANNs were necessary for the recognition of M partial

discharge sources. For the first case, two types of ANNs were developed, the first had only two outputs (two partial discharge sources) while the second had 12 outputs (12 partial discharge sources). In the case of only two outputs, the back propagation network, Kohonen self organising map and learning vector quantization network classified correctly the test patterns taken from the set of patterns used for training. However testing with unknown patterns resulted in a number of misclassifications. In the case of 12 outputs, the back propagation network provided satisfactory results for all the studied partial discharge sources compared to the other types of ANNs. When using M ANNs for M individual partial discharge sources, each network was trained to give unity output for the presence of a specific partial discharge source and zeros for the other partial discharge sources which called the counter sources. However, it was observed that different outputs were obtained for the same test partial discharge source when the network was trained with two different counter partial discharge sources.

Another approach which was not based on these distributions, but focussed on features that describe the shape of the partial discharge pulses, namely the apparent charge, rise time, fall time, width and area of the partial discharge pulse, had also been used [Mazroua *et al*, 1993]. These five features cooperate together to form the partial discharge pattern that were fed to a multilayer ANN trained using the back propagation algorithm. Artificially created cylindrical cavities of 2 *mm* in diameter and having depths ranging between 1.0 and 2.0 *mm* formed in acrylic disk specimens, were employed to test the capability of the ANN to discriminate between different partial discharge pulse patterns. Two different structures of ANNs were adopted to perform the classification task between three outputs (three depths). The first one with single hidden layer and the second one with two hidden layers. It was

observed that the two structures yielded the same results with 75% accuracy for the training patterns when used for testing whereas their discrimination performance decreased to 50% when the network was tested for 10 patterns which were not used for the training. When the difference between the cavity depth sizes under investigation increased, the classification accuracy increased to 95% and 70% for the training patterns and the patterns which were not used in the training. The ANN succeeded in discriminating between the patterns of the metallic electrode cavities and the cavities between the dielectric and electrodes with a success rate of 100% for both the training and the test data. The same features and the same ANN had been extended for the recognition of discharge sources of different types such as cavities and electric trees within the insulation system as well as the recognition of changes in the partial discharge shapes that were associated with the deterioration of ageing effects within the defects undergoing discharge [Mazroua *et al*, 1995]. A new feature given by the product of the actual test voltage and the apparent charge as well as the width of the pulse had been replaced by the multiplication of the pulse width by the apparent charge which resulted in six features. These were used to learn three different ANNs [Mazroua *et al*, 1994]; namely, the multilayer perceptron, nearest neighbour classifier and the linear vector quantization. It was noticed that the recognition capabilities of the three ANNs were comparatively equivalent.

The back propagation feed forward ANN had also been used to discriminate the partial discharge of three kinds of electrode systems which resulted in 24 different classes [Okamoto *et al*, 1995]. These groups were a needle-plane electrode system group, a spherical void electrode system group, and a crack void electrode system group. The input patterns based on the ϕ–q-n were divided into 20 phase windows and 16 magnitude

windows. Five patterns were used to represent each class. In order to examine the recognition ability of the network, two kinds of input pattern groups were used. One was a set of unlearned patterns and the other was a combination pattern of two different electrode groups. While the ANN could recognise successfully the patterns which were used for training, the classification accuracy reduced for the unlearned patterns to around 85%. In the case of combination of two electrodes, there was no output other than A, B output when the input patterns were a combination of an A electrode patterns and a B electrode patterns.

1.6.4.2 Conventional classifiers

There are number of conventional classifiers available for classification [Kreuger *et al*, 1993]:

- Minimum distance classifier
- Rate of recognition classifier
- Centaur score classifier

In the above methods, parameters such as the mean value, standard deviations of the features for different effects, are calculated. Then the distance between a finger print of unknown origin and the known defect is calculated to assign the finger print to the proper defect.

1.6.4.2.1 Minimum distance classifier

The minimum distance (Euclidean distance) classifier computes the distance from the unknown pattern to the mean of each class, and assigns the unknown pattern to the class to

which it is closest. The distance is given by [Tou and Gonzalez, 1974]:

$$d = \sqrt{(x_i - \mu_{ik})^t (x_i - \mu_{ik})} \qquad (1.11)$$

Where,

x_i is the i^{th} component of the classification features vector (finger print)

μ_{ik} is the i^{th} component of the mean values of the features vector belonging to a reference defect class k.

The disadvantage of this approach is that the classifier can not provide "I do not know" answers. So, a finger print which does not belong to any of the known problems may be misclassified as one of the known problems.

1.6.4.2.2 Recognition rate

The recognition rate is determined in the following steps [Kreuger *et al*, 1993]:

- Several samples of known defects are taken and measured for discharge.
- For each parameter the mean value is determined, and the standard deviation is derived from the scatter of these values.
- Now, each parameter has a most probable value between two limits (mean value – standard deviation and mean value + standard deviation).
- The unknown defect is measured and the value of each operator compared to that of the known defect.
- If the difference is small, a hit is recorded.
- In this way, all the parameters are compared and the number of results that coincide are recorded, this total number is called as 'recognition rate'. Actually,

recognition rate and Centaur score do not differ much in their discriminating power.

1.6.4.2.3 Centaur Score

The Centaur method creates percentage contour in the form of hyrerellipsoids (ellipse in 2-d space) around the mean value of features for a particular defect [Kreuger *et al*, 1993]. The size and shape of hyrerellipsoid is determined by the standard deviation of each feature and mutual correlation of the features. The Centaur method has been applied successfully for the recognition of created defects in insulation and in actual high voltage equipment. The use of the Centaur score method is, however, restricted to normally distributed data of a particular defect. By the careful design of the data base for discharge recognition e.g. by splitting a non normal distribution into several normal ones, this condition can be fulfilled in such a way that the misclassification can be avoided.

1.7 Partial Discharge Training Data

A carefully designed data base for partial discharge classification should produce high level of similarity between a finger print to be classified and a known insulation defect in the case of a correct recognition and a low level of similarity in all other cases. Several questions arise when creating the database. For example:

- How many features are sufficient for recognition?
- How many patterns of one partial discharge source are required for successful feature classification?

Usually as many features as possible are extracted from the patterns. However, the higher the number of features, the longer time is required to calculate the features and their classification. Some features may be useless for pattern recognition because they have no discriminating power. A compromise between the two conflicting requirements can be achieved by selecting only the best features by maximising certain criteria.

1.8 Motivation and Thesis Objectives

It has been observed that the partial discharge generated by ac voltage are significantly influenced by memory associated with the charge developed by preceding discharge events [Van Brunt *et al,* 1989]. Also it is well known that partial discharge phenomena generally involves stochastic processes in the sense that when a partial discharge pulse is generated, it leaves behind an imprint that can influence development of the subsequent discharge pulse. These imprints may, for example, be in the form of a negative ion space charge moving in the gas, molecules in excited metastable state or charges on the surface. These conclusions have been drawn from the study of various conditional and unconditional pulse-height and pulse-time-separation distributions [Van Brunt *et al*, 1989, 1991, 1993]. It is assumed that the discharge phenomenon consists of a sequence of pulses denoted by the pulse amplitudes q_n and time separation between successive pulses Δt_n. The time separation is measured between the n^{th} and the $(n+1)^{th}$ pulses. The possible conditional pulse-height and time-interval probability distribution functions may be of zero, first or second order. The zero order functions are obviously unconditional. For example, $p_0(q_n)dq_n$ is the probability that any discharge pulse occurs with magnitude in the range q_n to $q_n + dq_n$ is independent of n. The distributions p_1 and p_2 are conditional, e.g. $p_1(q_n/\Delta t_{n-1})dq_n$ is the

probability that the n^{th} discharge pulse for arbitrary n will have an amplitude between q_n to $q_n + dq_n$, when its time separation from the preceding pulse is restricted to the value Δt_{n-1}, and $p_2(q_n/\Delta t_{n-1})d\, q_n, q_{n-1})\, dq_n$ is the same, with the added restriction that the amplitude of the preceding pulse is also fixed to at a value q_{n-1}. An immediate indication that q_n and Δt_n are not independent can be drawn from the observation that a measured conditional probability is not equal to the corresponding unconditional distribution. Also, it is found that the conditional distribution changes its shape and position for different values of Δt_n. Therefore, it is the nature of the phenomenon that a significant correlation exists between successive discharge pulses. Hence, one can not assume that pulse amplitude and time separation from the previous pulse are independent random variables.

It is clear from the above that the pulse height phase resolved distributions contain no information about the correlation between successive pulses in the same half cycle and therefore provide no information about memory propagation effect. Therefore, the main aim of this thesis is to explore another effective technique which can take the memory effect into consideration and eliminate the averaging effect of these distributions as well as to get an overview of every single partial discharge pulse of the whole measurement. The texture analysis has been successfully applied to the image processing and remote sensing fields. However, it is yet to be tried for the partial discharge pattern recognition. In the present work, an attempt has been made to apply the texture analysis algorithm to discriminate different types of partial discharge sources. The main aims of the thesis are to attempt the following:

(a) Investigate the discriminating power of different texture analysis algorithms for

the partial discharge source classification.

(b) Compare the discriminating power of the texture analysis algorithms with the discriminating power of the (q-ϕ-n) distributions method for partial discharge sources classification.

(c) Determine the relative discriminating power of each feature individually.

(d) Determine the minimum number of features used at a time to achieve a good classification accuracy.

(e) Determine the best features selection technique to minimise the number of features which can be used to classify different partial discharge sources.

1.9 Thesis Organisation

The work reported in this thesis is organised in nine chapters:

The present chapter introduces the partial discharge phenomenon, reviews the conventional methods for partial discharge classification, limitations of these methods and briefly outlines the work carried out in this thesis.

Chapter 2 presents theoretical background of four different texture analysis algorithms. The first three of the algorithms are in the spatial domain while the last one in the frequency domain. These algorithms include the spatial gray level dependence method (SGLDM), the gray level difference histogram method (GLDHM), the gray level run length method (GLRLM), and the power spectral method (PSM), respectively.

Chapter 3 describes the experimental set up used for partial discharge measurements and the samples used to simulate the partial discharge sources. The standard partial discharge sources, which have been used in this study to evaluate the discriminating power of the texture analysis algorithms, include glow corona, streamer corona, surface discharge, internal discharge, single protrusion and multi protrusions which were created in the laboratory.

Chapter 4 utilises the minimum distance classifier for partial discharge classification using the features based on the texture analysis algorithms. The classification accuracy of each algorithm has been investigated and the classification accuracy of each feature individually as well as for the combination of two features at a time have been studied.

Chapter 5 presents the application of a direct feature selection technique in the form of transformed divergence analysis. Transformed divergence has been used to measure the distance between the classes in the feature space. The optimum features have been selected according to the maximum separation between the different partial discharge classes.

Chapter 6 investigates the use of a non parametric classifier in the form of back propagation artificial neural network (ANN). The application of ANN were mainly motivated by the observation that some of the partial discharge sources are not normally distributed. Thus the assumption for minimum distance (mahalanobis distance) and the transformed divergence were not satisfied. A three layer feed forward ANN has been developed to classify different partial discharge sources.

After the discriminating power of texture analysis algorithms is established, in chapter 7, the GLDHM features have been used to distinguish two partial discharge sources generated from high voltage cable as a practical example. A paper insulated lead covered belted cable has been used to generate internal discharge, which take place in the microvoids within the insulating paper. Surface discharge at the end terminal of the cable are also produced under unprevented conditions.

Chapter 8 describes the application of an indirect method has been used to reduce the number of features by mapping the original set of features into a new set of features lesser in number in which the separation between the classes are maximum . The principal component transformation of the four texture analysis algorithms features have been investigated to determine the classification accuracy of the main principal components.

Chapter 9 presents summary of main conclusions of the thesis and includes few recommendations for further research work.

Chapter 2

Basics of Texture Analysis

2.1 Introduction

From the discussion in chapter-1, it is observed that the effect of the memory propagation between the successive partial discharge pulses are not taken into consideration by the phase resolved pulse height analysis. The memory propagation means the dependence of the magnitude of particular pulse on the magnitude of the preceding pulse and the time separation between the two pulses. Texture analysis algorithms mainly investigate the gray level variation in images [Haralick *et al,* 1973]. In other words, they investigate the relative magnitude of gray level and its spatial distribution in images. Therefore, texture analysis algorithms are expected to create improved discriminating features between different partial discharge sources by investigating the relationship between the adjacent partial discharge pulses. The memory propagation between successive pulses can be considered in the form of relative pulses magnitude as well as the spatial distribution of these pulses.

Since the application of the texture analysis to the partial discharge sources classification is new, the objective of this chapter is to develop an understanding of the terminology being used and conceptual background of texture analysis. A careful review of the theory of texture is presented in this chapter.

2.2 Background

Application of texture analysis has always been an interdisciplinary field of endeavor. It involved people from remote sensing application, medical application, metallurgy, computer science, and military application. In remote sensing application, the texture analysis has been used for land use studies, terrain classification, cloud analysis, and sea ice detection. In medical science, it has found applications in chest, dental, and other x-ray pictures and chromosomal analysis. In metallurgy, it has been used to study electron micrographs of various metal and alloy samples. In computer science it has been used for texture-based techniques for efficient automation. Military applications use texture information for the detection of the targets. In fact, it is an extremely difficult task to list all possible uses of texture analysis because texture has been a part of our physical impression of the visual world. The following sections outlines, in brief, the concepts and developments in texture analysis.

2.3 Concepts in Texture Analysis

The information content of an image can be defined in terms of three fundamental features: spectral tone, texture, and context. Spectral features describe tonal variations in spectral bands of the electromagnetic spectrum. Texture features contain information about the spatial distribution of spectral variations, and contextual features provide adjacency information in an image area. Haralick *et al* (1973) suggested that textural and spectral properties are present in an image simultaneously but under a given condition, one property can dominate the other. When a small region has a wide variation in gray level, the dominant property is texture for that area. If the area has very little variation in gray

values, tone becomes the main property. In other words, as the number of distinguishable discrete tonal features decrease, the tonal properties dominate, whereas textural properties dominate with the increase in distinguishable discrete tonal features.

2.3.1 The notion of texture

Hawkins (1970) suggested that the notion of texture depends on three things:

(i) Repetition of some local order over a region which is large compared to the order's size.

(ii) Non-random arrangement of elementary parts in the local order

(iii) Elementary parts of the local order being roughly uniform entities and having approximately similar dimensions everywhere within the textured region

2.3.2 Terms and definitions in texture analysis

It has been extremely difficult task for researchers to define the meaning of texture. Everyone understand the concept of texture but it is extremely difficult to define the term texture. Generally, researchers have defined it qualitatively and have tried to use numerical methods based on the quality of the texture they want to extract. Qualitatively, texture is described by various terms like coarseness, contrast, directionality linelikeness, regularity and roughness (Tamura *et al,* 1978). Coarseness depends upon the pattern and size of texture elements. Contrast depends upon the ratio of black and white area, sharpness of edges and period of repeating patterns. Directionality depends upon the shape of the elements and placement rules. Linelikeness refers to the shape of the primitives of the

textures, supplements coarseness, contrast, and directionality. Regularity relates to the variation in placement rules of the primitives or texture elements. If the placement rules are fixed and there is less variation in tone, the texture is said to be regular and coarse. But if the placement rules for texture elements are less fixed and there is more variation in gray levels, the texture becomes irregular and fine. Roughness is a mental concept related to real world of three dimensional objects.

Apart from these descriptions, texture is essentially a resolution dependent phenomena. It implies that a change in the scale of imagery will change the whole perception of coarseness, contrast, and directionality of the texture. Thus a texture which is fine at a particular resolution may appear coarse with a change in the resolution of the imagery. In totality, image texture can be described as an area phenomenon rather than a point consisting of many texture elements which can be statistically described with some placement rules. In nature, spatial organization of these texture elements gives rise to deterministic as well as random textures. The deterministic texture is one in which the texture elements can be strictly represented by some fixed placement rules. A random texture is one in which the texture elements follow some statistical laws and are stochastic in nature. With the above background, it is possible to introduce some definitions of texture:

- Texture relates to the relationships between gray levels in neighboring resolution cell which contribute to the overall appearance or visual characteristics of an image (Thomas, 1977).

- Texture could be defined as a structure composed of a large number of more or less ordered similar elements or patterns without one of these drawing special attention so that a global unitary impression is offered to the observer (Gool *et al*, 1985).

- Visual texture refers to the impression of roughness or smoothness created by tone or repetition of visual pattern across the surface (Irons and Peterson, 1981).

- Texture is simply a repeated variation in tone over relatively small areas (Swain and Davis, 1978).

- Texture is the visual effect which is produced by spatial distribution of tonal variations over relatively small area (Andera Baraldi and Flavio Parmiggiani, 1995).

Apart from these definitions of texture, there are some other terms which are quite frequently used in texture investigations. Definitions of these terms are given below:

2.3.2.1 Texture analysis

It is the process in which image gray level spatial relationships concerned with shape position and periodicity are assigned mathematical descriptors suitable for characterizing the properties in such a way that they may serve as features for extraction in machine classification (Thomas, 1977).

2.3.2.2 Texture feature

Relationships including the statistical distribution of gray levels or information about the

boundaries of edges arising from gray level gradients or discontinuities when represented by mathematical or information theory measures are called texture features. These are vectors which contain a large amount of information related to the discrimination between class types in a given set of classes.

2.4 Texture Analysis and Computational Methods

In attempt to simulate the process of visual perception for texture analysis, researchers have tried to extract texture by capturing some form of the first- and second-order statistics. In computational texture analysis, the main problems are[Ehrich and Foith, 1978]:

(i) Identification of the class label for a given textured region;

(ii) Description of a given textured region;

(iii) Construction of boundaries between the major texture regions.

The first problem is the texture classification, which requires derivation of a set of rules to discriminate between a given finite number of texture classes. This problem is usually approached by using some statistical methods for learning the characteristic parameters of each defined class. Description of textured region is more difficult as it deals with the extremely complex structure of texture. Accurate description of texture requires an understanding of the structural dependencies among basic and composite textural elements. Boundary definition is the most difficult problem because texture properties may undergo slow spatial variation that is difficult to detect or quantify.

2.5 Developments in Texture Classification

From classification point of view, texture has been studied at two levels: statistical and structural levels [Haralick, 1979]. The statistical approach aims at a global characterization of texture. Statistical properties of the spatial distribution of gray levels are used as texture descriptors. The key feature of this approach is that the description solely depends upon point properties, with no explicit use of element or subregions. At this level, texture is assumed to be defined by a set of statistics derived from a large ensemble of local picture properties. The second level of structural analysis attempts to isolate the primitives or units of texture and describe the relations between these in the texture pattern.

The statistical and structural approaches have their own advantages. A lot of work has been done as far as statistical approaches are concerned since they are simple. Structural approach, though apparently realistic, find their utility in cases where it is possible to detect some identifiable patterns. This does not mean that they can not be applied to other situations. The problem is that, with an increase in the scene complexity, computational loads go beyond feasible limits for application of structural approaches. Hence, such approaches have been limited to the analysis of well defined structures only.

2.5.1 Statistical approaches to texture analysis

Statistical approaches are based on deriving a set of statistics of a particular order in the spatial or frequency domain and characterizing texture completely by these statistics. Statistical approaches can be divided into spatial domain analysis and frequency domain analysis.

2.5.1.1 Spatial domain analysis

This approach can be further divided into first order statistics and second order statistics.

(a) First order statistics

These include measures based on probability distributions of single pixel attributes. The most common of these are the mean and the standard deviation. The mean is a measure of overall image brightness, whereas the standard deviation is a measure of variability in spectral values within the image. Other measures include the coefficient of variation, higher moments of different order, spatial autocorrelation, gray level difference and run length statistics. In one of the earliest articles on computational texture analysis, Hawkins (1970) classified texture measures into four groups based on properties such as spatial frequency, gray level, local shape, and higher orders. The spatial frequency approach was based on the fact that most objects in nature have an identifiable spatial frequency of variation. Gray level measures included features like the average, variance, histogram, and gray level entropy.

Galloway (1975) suggested that texture could be measured by forming a gray level run length matrix (GLRLM) by counting the number of gray level runs of various lengths and then calculating various statistical features derived from this matrix that could be input to the classifier.

Conners and Harlow (1980) proposed the gray level difference histogram measure (GLDHM) in which a probability distribution function for the absolute difference in the

gray levels of pixels, separated by a fixed spatial relation, is constructed. From this distribution, various features are calculated to classify texture.

(b) Second order statistics

These methods are based on the spatial arrangement of adjacent pixels and consider the second order probability distribution. Since it is possible to derive the first order measures from the second order measures, the first order measures provide less information than second order measures.

Haralick *et al* (1973) presented one of the most widely used approaches to texture analysis. It was based on the estimation of the second order joint probability density functions. The approach is called as the spatial gray level dependence matrix (SGLDM) approach. The intermediate result of this approach is in the form of so called co-occurrence matrices. From this distribution, various features are calculated to classify texture.

2.5.1.2 Frequency domain analysis

This approach is based on using the power spectrum method (PSM). It is important to explain the concept of spatial frequency. Spatial frequency is the image analog of the frequency of a signal in time. A sinusoidal signal with high frequency alternates rapidly, whereas a low frequency signal changes slowly with time. Similarly, an image with high spatial frequency, say in the horizontal direction, exhibits frequent changes of brightness with position horizontally. A picture of a crowd of people would be a particular example. Typically, an image is composed of a collection of both horizontal and vertical spatial

frequency components of different strengths and these are what the discrete Fourier transform indicates. Entries in the Fourier transformed image $\phi(r,s)$ represent the composition of the original image in terms of the spatial frequency components, both vertically and horizontally. The upper left hand pixel in $\phi(r,s)$ i.e. $\phi(0,0)$ is the average brightness of the image. This is the component in the spectrum with zero frequency in both directions. Thereafter pixels of $\phi(r,s)$ both horizontally and vertically represent components with frequencies that increment by 1/k, where the original image is of size k×k pixels. The features commonly used with the PSM are (Weszka *et al,* 1976; Conners and Harlow, 1980; Gool *et al,* 1985):

i- Annular-ring sample geometry.

ii- Wedge sampling geometry.

iii- Parallel slit sampling geometry.

2.5.2 Structural approaches to the texture analysis

These approaches consider texture as an arrangement of a set of spatial sub-patterns according to certain placement rules. The sub-patterns themselves are, in general, made up of smaller sub-patterns, positioned according to some placement rules. This recursive nature of the approach captures the hierarchical structure of natural scenes where both the sub-patterns and their placement may be characterized statistically. The approach is primarily based on the concepts of unit pattern and well defined placement rules. The following steps form the key to this approach (Haralick, 1979):

- Location of unit patterns primitives;
- Extraction of features that characterize primitives;
- Extraction of features that characterize the placement rules of unit patterns.

The term primitive is used to represent a set of connected resolution cells which is characterized by a list of attributes. The simplest primitive can be a single pixel with its gray level as its attribute. Thus, primitives are maximally connected set of resolution cell having some specific properties. Various attributes can be used to characterize a primitive including average gray level, size or number of connected pixels, elongation, spread, and orientation of primitives axes.

2.6 Statistical Texture Analysis Algorithms

Consider that the image to be analyzed is rectangular and has *Nx* resolution cells in the horizontal direction and *Ny* resolution cells in the vertical direction and the gray level appearing in each resolution cell is quantized to *Ng* levels. For example, consider a 4×4 image with four gray levels 0-3 such as [Haralick 1979]

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Each resolution cell in the image, excluding those on the periphery of the image, are having eight nearest neighbor cells. In the following sections different texture analysis algorithms which can be used to extract the discriminating information, for the above

example, are described.

2.6.1 Spatial gray level dependence method (Co-occurrence)

Gray level co-occurrence matrix describes the frequency of one gray tone appearing in a spatial relationship with another gray tone, within the area under investigation. In other words, the image can be specified by a matrix of relative frequency P(i,*j*) with which two neighboring resolution cells separated by distance *d* occur on the image, one with gray tone *i* and the other with gray tone *j*. Such matrices of gray tone spatial dependence frequencies are a function of the angular relationship as well as a function of the distance between them. The following table shows the general form of any co-occurrence matrix. For example, the element in the (2,1) position is the number of times two gray levels of values 2 and 1 occurred adjacent to each other in a given direction.

Gray level	Gray level			
	0	1	2	3
0	0,0	0,1	0,2	0,3
1	1,0	1,1	1,2	1,3
2	2,0	2,1	2,2	2,3
3	3,0	3,1	3,2	3,3

When the relationship is nearest horizontal neighbor, there will be *2(Nx-1)* neighboring resolution cell pairs on each row and there are *Ny* rows providing a total of *2Ny(Nx-1)* nearest horizontal neighbor pair. For the vertical direction, there will be *2Nx(Ny-1)* nearest vertical neighbor pair. For the left and right diagonal, there will be *2(Ny-1)(Nx-1)* nearest vertical neighbor pair. For the above example, the co-occurrence matrices in the horizontal direction, vertical direction, left diagonal direction, and right diagonal directions for

where $p(i,j)$ are the entries of the co-occurrence matrices

Entropy (ENT), is a measure of the degree of complexity or heterogeneity or disorder in the image. If all the values of gray level are equally probable then the entropy will be high. This will be true for very busy texture. It will be small for homogeneous images i.e. when entries in co-occurrence matrix are very different. The entropy is a measure of randomness in the image, and can be defined as,

$$ENT = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p(i,j)\log(p(i,j)) \qquad (2.2)$$

Inertia or **Contrast (CON)**, is a measure of local variation in an image and calculates the moment of co-occurrence matrix about its mean diagonal. It is defined as,

$$CON = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} (i-j)^2 p(i,j) \qquad (2.3)$$

Inverse difference moment(IDM), is a measure of local similarity in the image. It characterizes the image in terms of lack of variability in the gray level value. IDM is high when diagonal concentration of element is high. IDM is defined as,

$$IDM = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \frac{p(i,j)}{1+(i-j)^2} \qquad (2.4)$$

Correlation (CORR), is a measurement is expected to be associated with linear structures in the image. It is a measure of linear gray level dependence and defined as,

$$CORR = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \frac{p(i,j)(i-\mu_x)(j-\mu_y)}{\sigma_x\sigma_y} \qquad (2.5)$$

Variance (VAR), is a measure of heterogeneity of the image.

$$VAR = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(i-\mu_x)(j-\mu_y)p(i,j) \qquad (2.6)$$

2.6.2 Gray level difference histogram method (GLDHM)

In the gray level difference histogram method, the probability distribution function of the absolute difference in the gray level of pixels, separated by a fixed spatial relation, is constructed. From this method, one gets four vectors in the four directions. Each vector has length equal to the difference between the maximum and minimum gray level in the image plus 1. For the example given in section 2.6, the differences could be 0, 1, 2 and 3. Zero means that the two adjacent cells have the same gray level and three means that one of the two adjacent cells has the maximum gray level value and the other has the minimum gray level value. The gray level difference histogram for the distance equal to one pixel in the different directions can be given as following:

Differences	0	1	2	3
Frequency	16	6	2	0

Horizontal

Differences	0	1	2	3
Frequency	12	8	4	0

Vertical

Differences	0	1	2	3
Frequency	4	8	6	0

Left diagonal

Differences	0	1	2	3
Frequency	10	8	0	0

Right diagonal

The length of the histograms depends upon the gray level of the various pixels in the image only, therefore their lengths are the same in the different directions. From these histograms, various textural features can be calculated to classify texture. These features are mean, angular second moment (ASM), contrast (CON), entropy (ENT) and inverse difference

moment (IDM) as defined below.

$$MEAN = \sum_{i=0}^{n-1} p(i)\, i \tag{2.7}$$

$$ASM = \sum_{i=0}^{n-1} (p(i))^2 \tag{2.8}$$

$$CON = \sum_{i=0}^{n-1} p(i)\, i^2 \tag{2.9}$$

$$ENT = -\sum_{i=0}^{n-1} p(i) \log(p(i)) \tag{2.10}$$

$$IDM = \sum_{i=0}^{n-1} \frac{p(i)}{1+i^2} \tag{2.11}$$

where $p(i)$ are the entries of the gray level difference histogram.

2.6.3 Gray level run length method (GLRLM)

GLRLM extracts gray level run which is a set of consecutive image points having the same gray level. The run length matrix in a given direction will be of the form:

Gray level (a)	Run length (b)			
	1	2	3	4
0	0,1	0,2	0,3	0,4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4

The entry $p(i,j)$ in a run length matrix specifies the number of times the image contains a run of length j in the given direction consisting of points having gray level i. For example, the element in the (2,3) position shows the number of times that three adjacent cells have a

gray level value of 2 in a given direction. In this method for a given direction, we get a matrix which has dimension $a \times b$, where a correspond to the values varying from minimum to maximum gray level in the image and b equals the number of cells of the original image in the corresponding direction. For the example used earlier the run length matrix in the different directions will be as following:

1	2	0	0
0	2	0	0
0	1	1	0
0	1	0	0

Horizontal

0	1	1	0
0	2	0	0
3	1	0	0
2	0	0	0

Vertical

3	1	0	0
2	1	0	0
5	0	0	0
2	0	0	0

Left diagonal

1	2	0	0
2	1	0	0
1	2	0	0
2	0	0	0

Right diagonal

For each one of the above matrices, various features can be calculated in the different directions. Following notations have been used for calculating the run length features:

n_g = number of gray levels in the image

n_r = number of different run lengths that occur

p = number of pixels in the image

$p(i,j)$ are the entries of the run length matrices

$$R = \sum_{i=0}^{n_g-1} \sum_{j=1}^{n_r} p(i,j)$$

This technique extracts the following five textural features.

Short run emphasis(SRE): This feature emphasizes short runs. In case of lot of gray

level variation and hence runs of small length, this feature will assume high value compared to images which will have less variation and many runs of longer lengths. It is defined as,

$$SRE = \sum_{i=0}^{n_g-1}\sum_{j=1}^{n_r} p(i,j) / j^2 / R \qquad (2.12)$$

Long run length emphasis(LRE): This feature emphasizes long runs and hence images of less variation. It can be defined as,

$$LRE = \sum_{i=0}^{n_1-1}\sum_{j=1}^{n_r} p(i,j) j^2 / R \qquad (2.13)$$

Gray level non-uniformity (GLN): This feature measures the gray level non-uniformity. When runs are equally distributed throughout the gray levels, the function takes its lowest value. It is defined as,

$$GLN = \sum_{i=0}^{n_g-1}\left[\sum_{j=1}^{n_r} p(i,j)\right]^2 / R \qquad (2.14)$$

Run length non-uniformity (RLN): This function measures the non-uniformity of the run lengths. If the runs are equally distributed throughout the lengths, the function will have a low value. It can be defined as,

$$RLN = \sum_{j=1}^{n_r}\left[\sum_{i=0}^{n_g-1} p(i,j)\right]^2 / R \qquad (2.15)$$

Run percentage (RP): This function is the ratio of total number of runs to the total number of possible runs if all had length of one. It has its lowest value for image with most linear structure.

$$RP = \sum_{i=0}^{n_g-1}\sum_{j=1}^{n_r} p(i,j) / p \qquad (2.16)$$

2.6.4 Fourier power spectrum method

The Fourier transform of a two dimensional function $f(x,y)$ is defined as [Pratt, 1978],

$$\phi(r,s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-2\pi j(rx+sy)} f(x,y)dxdy \qquad (2.17)$$

and the Fourier power spectrum is given as $|F|^2 = FF^*$ (where * denotes the complex conjugate). In case of a coarse texture, which has less gray level variation, the value of Fourier power spectrum will be high and concentrated at the origin. While in a fine texture, which has more gray level variation, this value will be more spread out. Thus, to analyze any texture, a set of features that should be useful are the averages of Fourier power spectrum over ring-shaped regions or vertical and horizontal slits [Pratt, 1978].

For a $n \times n$ digital images, consider that $f(i,j)$ is the gray level of a pixel at location (i,j). The Fourier transform of the image, in discrete form, is defined as,

$$\phi(r,s) = \frac{1}{n^2}\sum_{i,j=0}^{n-1} f(i,j)e^{-2\pi j(ir+js)} \qquad (2.18)$$

The calculation of the discrete Fourier transform (DFT) is very much time consuming [Richards, 1993]. Therefore, in this work, the fast Fourier transform (FFT) has been used. In FFT, the computation time increases almost linearly with the number of inputs, whereas in case of the DFT the time increases quadratically. Application of the FFT requires that the number of inputs, for example k, to be continuously power of 2 (say $k=2^m$). Since the images used have two dimensions, the FFT is required to be calculated for each dimension individually (i.e. it is necessary to transform each row individually to generate an intermediate image, and then transform this by columns to yield the final result). Thus, in this study, the dimensions of the images will have the power of 2. Also the application of the FFT requires that the order of the input data fed into the algorithm are rearranged for

odd and even inputs before the technique can be employed. For example, if we have 8 inputs with order 0,1,2,3,4,5,6,7 the final arrangement should be 0,4,2,6 and 1,5,3,7. This can be achieved simply by a process known as bit shifting. To do this, the index k of the input data X(k) is expressed in binary notation, the binary digit is reversed, and the new binary number is converted back to decimal form. For example;

$$\begin{array}{llllllll}
X(0) & \Rightarrow & X(000) & \Rightarrow & X(000) & \Rightarrow & X(0) \\
X(1) & \Rightarrow & X(001) & \Rightarrow & X(100) & \Rightarrow & X(4) \\
X(2) & \Rightarrow & X(010) & \Rightarrow & X(010) & \Rightarrow & X(2) \\
X(3) & \Rightarrow & X(011) & \Rightarrow & X(110) & \Rightarrow & X(6) \\
X(4) & \Rightarrow & X(100) & \Rightarrow & X(001) & \Rightarrow & X(1) \\
X(5) & \Rightarrow & X(101) & \Rightarrow & X(101) & \Rightarrow & X(5) \\
X(6) & \Rightarrow & X(110) & \Rightarrow & X(011) & \Rightarrow & X(3) \\
X(7) & \Rightarrow & X(111) & \Rightarrow & X(111) & \Rightarrow & X(7)
\end{array}$$

The standard set of texture features based on horizontal and vertical slits of the discrete Fourier power spectrum are of the form of,

Horizontal slits: $$S1(m) = \sum_{s(m)}^{s(m+\Delta m)} |\phi(r,s)|^2$$

Vertical slits: $$S2(m) = \sum_{r(m)}^{r(m+\Delta m)} |\phi(r,s)|^2$$

2.7 Partial Discharge and Texture Analysis

It is clear from the above description of the texture analysis techniques, used to extract discriminative features, that these can differentiate between different textures by using the gray level variation between the pixels. The partial discharge is an electrical phenomena which gives rise to electric pulses of stochastic nature. The measurement of partial discharge by using wideband detector is very useful to detect the partial discharge on pulse base as shown in Fig. 2.1. Fig. 2.1a shows a typical Glow corona pulses measured for the

positive half cycle of the applied voltage. The half cycle was divided into 2048 phase window. By using a low pass filter which makes each phase window represented by the average value of the neighborhood, one can get the low frequency component as shown in Fig. 2.1b. Now the filtered partial discharge pulses can be directly found by subtracting the low frequency component from the original measurement, as shown in Fig. 2.1c.

If more than one cycle (say m) could be measured and the measurements are arranged in array such that the number of horizontal inputs n is the number of windows in each cycle and the number of vertical inputs m is the number of the measured cycles. In other words, this array consists of m rows and n column. Figs. 2.2 and 2.3 are examples of different partial discharge measurements, shown like images. Each image is of 64X64 pixels (i.e.64 cycles and from each cycle only 64 phase windows are selected around the peak of the applied high voltage waveform, for clarity). If the gray level values, used in texture analysis algorithms, are replaced by the partial discharge pulses magnitude, the same concept of texture analysis algorithms can be used to study the partial discharge pulses magnitude variation. By studying the partial discharge magnitude variation, the relation between the adjacent pulses can be determined so that the effect of the memory propagation could be considered.

Whereas these pulses are stochastic in nature and it is very difficult to formulate any relationship between them, the statistical approach will be used to generate the features which can describe the surfaces of these images. Each image will be constructed for a particular partial discharge source i.e. there is no more than one partial discharge source in the same image. Hence, the problems of computational texture analysis like the texture description and the boundary between different texture regions are not considered in this

thesis. In this thesis, the attempt is to classify or identify the source of each image by studying the relationship between the pulses heights and their spatial distributions.

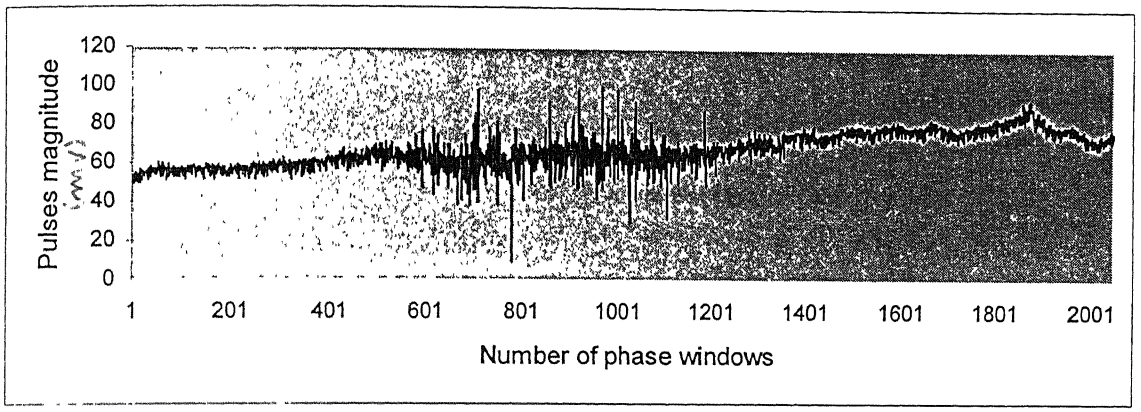


Fig. 2.1a: Original glow corona pulses

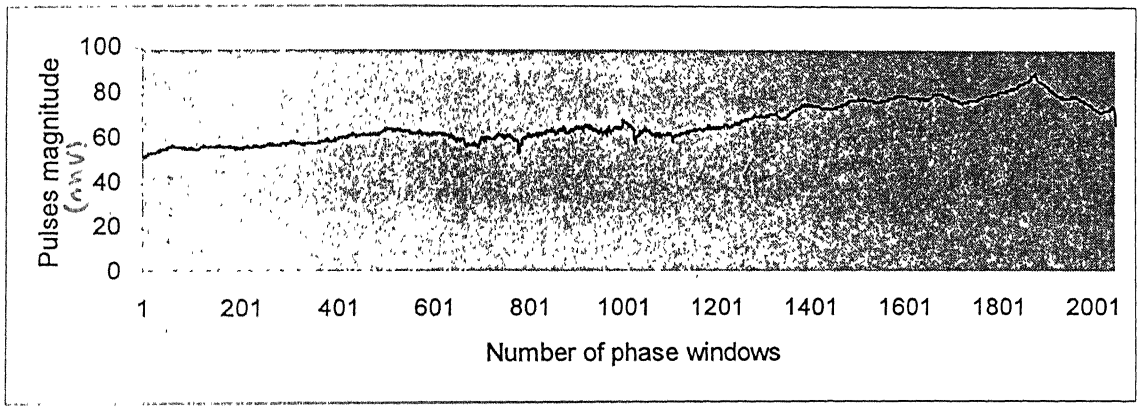


Fig. 2.1b: Low frequency component

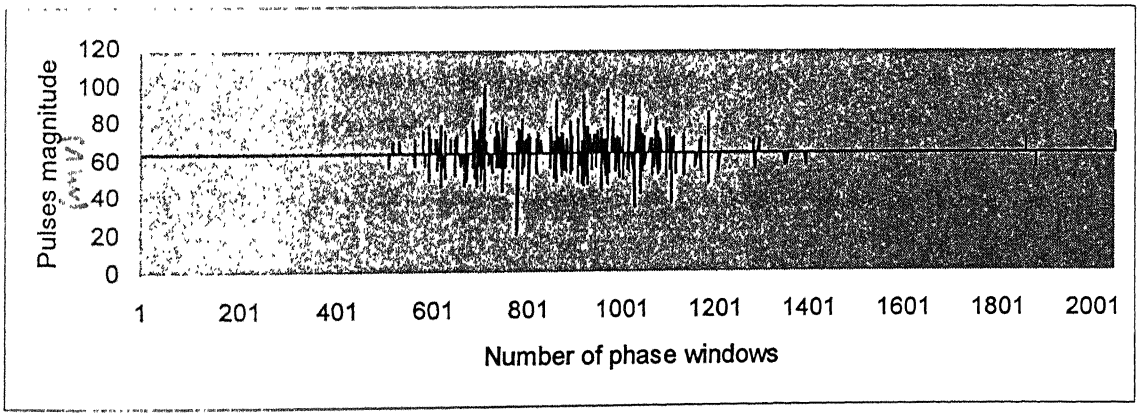


Fig. 2.1c: Filtered glow corona pulses

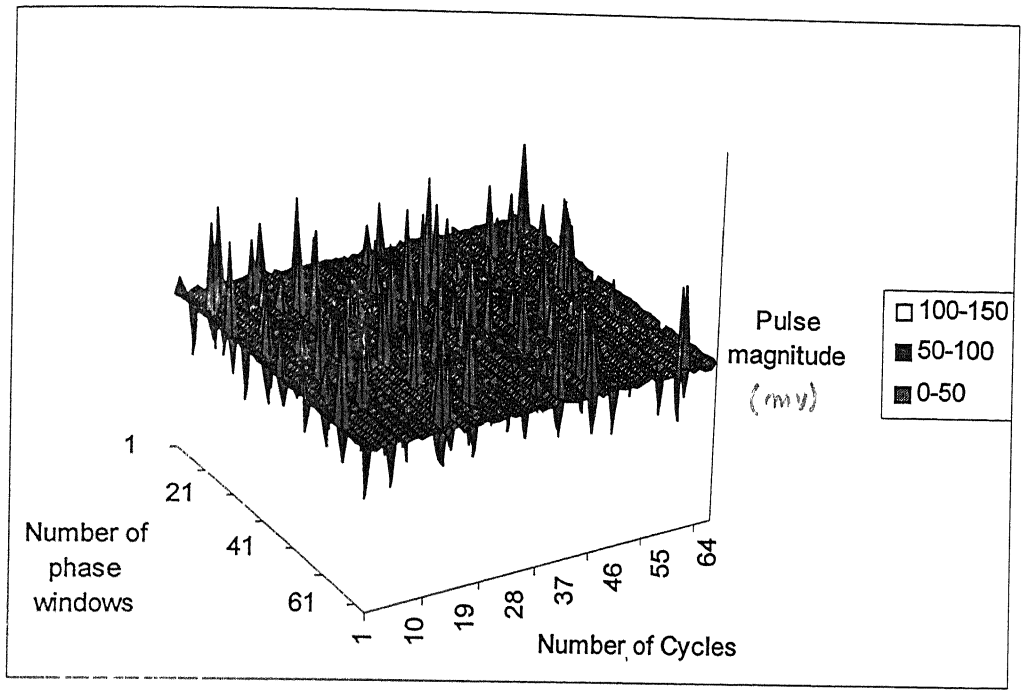


Fig. 2.2: 64X64 image for glow corona

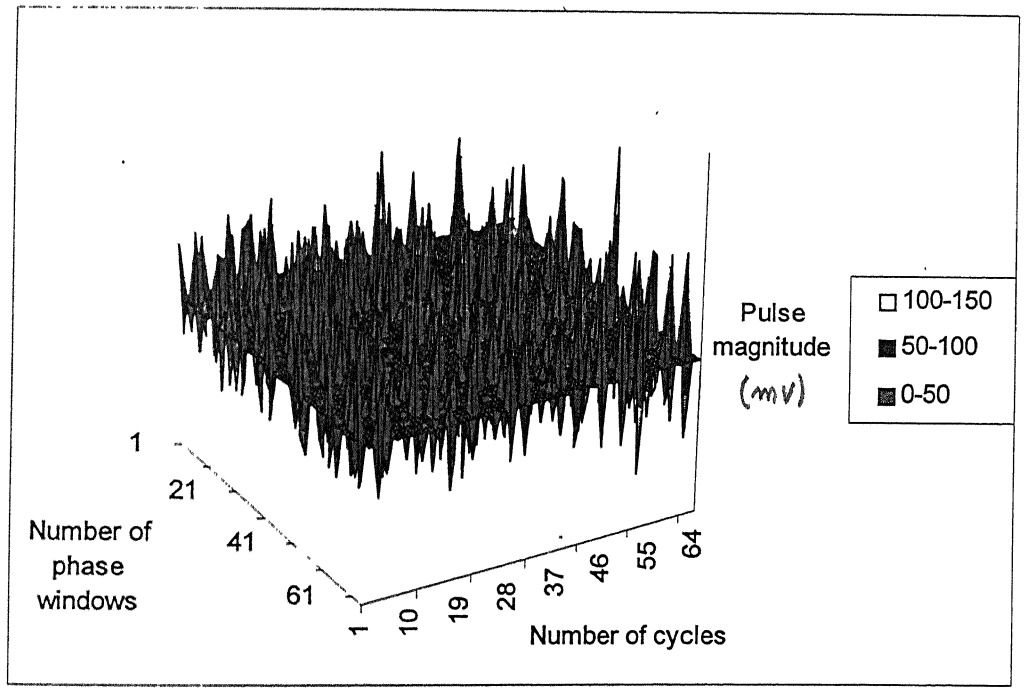


Fig. 2.3: 64X64 image for streamer corona

2.8 Conclusion

In this chapter, the concept of texture has been discussed and different approaches for texture analyses were described. Four different techniques have been explained in detail. Three techniques are in the spatial domain and one in the frequency domain. These techniques are the spatial dependence gray level method (SGLDM), gray level difference histogram method (GLDHM), gray level run length method (GLRLM) and the power spectrum method (PSM). The features generated from these techniques have been used in the subsequent chapters to discriminate between different partial discharge sources. The discriminating power of these four techniques as well as the relative discriminating power of the features of each technique will be investigated for the partial discharge sources classification.

Chapter 3

Experimental Set Up

3.1 Introduction

Generation of different partial discharge patterns is the first step to investigate the application of texture analysis algorithms in partial discharge source classification. To generate these patterns, different partial discharge sources have to be simulated in the laboratory and a suitable measuring circuit has to be used. This chapter explains how different types of partial discharge sources have been generated and measured in the laboratory. The description of the samples and equipment used as well as the circuit utilized for recording of the partial discharge patterns are presented in the present chapter.

3.2 Measurement Circuit

In practice, the most widely applied partial discharge measurement technique is electric pulse detection. This method is based on the measurement of the current impulses caused by a discharge in the defect. The partial discharges were measured using a straight detection circuit as shown in Fig. 3.1. This configuration permits the test sample to be grounded directly. The measurement set-up is principally composed of a high voltage source, test sample, coupling capacitor, measuring impedance, filter, amplifier, digital storage oscilloscope and personal computer. The high frequency partial discharge pulses

flow through the coupling capacitor and the detection impedance. The measured pulses are filtered, amplified, digitised by a digital storage oscilloscope and then processed by a computer. The specifications of all these components are given in the following sections.

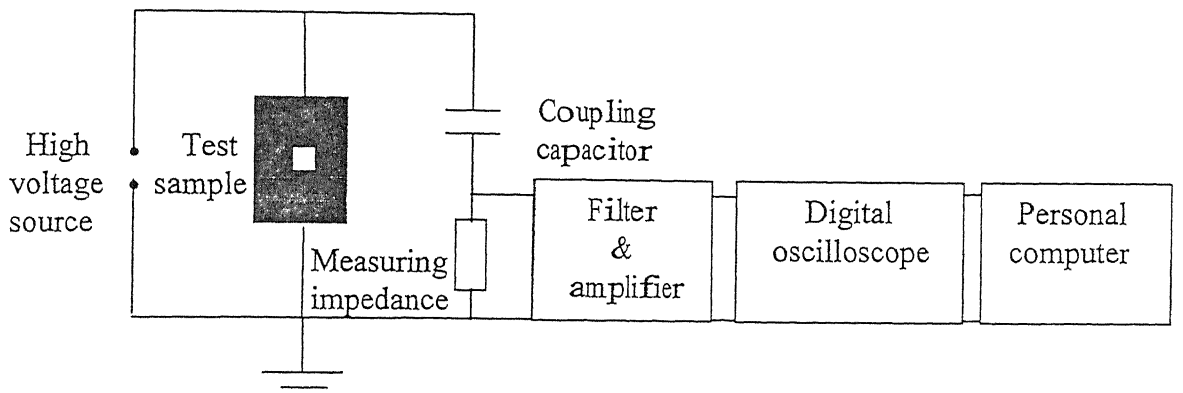


Fig. 3.1: Partial discharge detection circuit

3. 2.1 High voltage supply

The variable high voltage supply was obtained from a 100 kV, 50 kVA, ac power frequency partial discharge free test transformer. The high voltage terminal of the transformer was connected to a 1.1 nF, 100 kV partial discharge free high voltage coupling capacitor with the help of a 7.5 cm diameter aluminum pipe. The measurement of the partial discharge pulses were taken from the low voltage end of the coupling capacitor. Both the test transformer and the coupling capacitor high voltage terminals are provided with domes in order to prevent any partial discharge up to their rated voltages. These were verified as being partial discharge free up to 95 kV. The coupling capacitor was connected to the test sample with the help of a 3 cm diameter flexible pipe which has bell shape end to prevent the partial discharge on the connecting leads.

3.2.2 Measuring impedance

A *RLC* parallel circuit was used as the measuring impedance. It contains an inductance which is used to suppress the 50 Hz alternating current and its harmonics, such that the unbiased signal of partial discharge can be applied to the measuring circuit. With the help of a variable resistor in the range of 150Ω to 12 kΩ the measurement sensitivity can be adapted to the corresponding test requirement. Also it contains a spark gap, a charge eliminator and four diodes for over voltage protection. This impedance was also used to facilitate pulse observations since the inductance and capacitance produce an oscillatory response which persists much longer than the initiating pulse.

3.2.3 Filter and amplifier

Since the partial discharge pulses are , in general, a series of very short duration pulses having a magnitude of the order of millivolts super imposed on power frequency, these has to be filtered and amplified. An electric filter is often a frequency selection circuit that passes a specified band of frequencies and blocks or attenuates signals of frequencies outside this band. Filters may be classified in a number of ways [Ramakant, 1993]:

(i) Analog or digital

(ii) Passive or active

(iii) Audio or radio frequency.

Analog filters are designed to process analog signals. This is natural when dealing with signals that are continuous. However, digital filters process deals with discrete signals. Depending upon the type of the elements used in their construction, filters may be

classified as passive or active. Elements used in passive filters are resistors, capacitors and inductors. Active filters, on the other hand, employ transistors or operational amplifiers in addition to the resistors and the capacitors. The type of elements used dictates the operating frequency range of the filter. For example, RC filters are commonly used for audio or low frequency operation, whereas LC filters are employed at radio frequency or high frequency. In the audio frequency, inductors are often not used because they are very large, costly and may dissipate more power. An active filter offers the following advantages over the a passive one:

(i) Gain and frequency adjustment flexibility: Since the operational amplifier is capable of providing a gain, the input signal is not attenuated as in the case of passive filters. In addition, an active filter is easier to tune or adjust.

(ii) Reduced loading problem: Because of the high input and low output resistances of the operational amplifier, the active filters do not cause excessive loading to the source.

(iii) Cost effectiveness: Typically, active filters are more economical than passive filters. This is because of the availability of variety of cheaper operational amplifiers and the absence of inductors.

The most commonly used filters are low pass filter, high pass filter, band pass filter, band reject filter and all pass filter. In this study, since the measured partial discharge pulses have to be filtered from the low frequency harmonics and the very high frequency

interference, analog active wide band pass filter was used. A wide band pass filter can be found by simply cascading high pass and low pass sections. To obtain a ± 20 dB / decade band pass filter, first order high pass and low pass sections are cascaded. For a ± 40 dB / decade band pass filter, second order high pass and second order low pass sections are cascaded. In this study, first order high pass and low pass sections have been used. For the low pass filter, the higher cut off frequency was 400 kHz. Whereas, for the high pass filter, the lower cut off frequency was 40 kHz. The filter was constructed having the operational amplifier as the active element, capacitor and resistors as passive elements. The design of the low pass filter was done by selecting a capacitor C and then calculating the value of the resistance R with the relation;

$$R = \frac{1}{2\pi f_H C} \tag{3.1}$$

Where, f_H is the high cut off frequency. The schematic diagram of the low pass filter is shown in Fig. 3.2. The values of R_1 and R_F determine the pass band gain A_F as following:

$$A_F = 1 + \frac{R_F}{R_1} \tag{3.2}$$

For the low pass filter, the capacitor C was selected as $0.1nF$ and the resistor R as $0.4\ k\Omega$. The high pass filter was constructed simply by interchanging the position of the resistor R and the capacitor C used in the low pass filter. The schematic diagram of the high pass filter is shown in Fig. 3.3. The values of R and C for low cut off frequency of 40kHz were

selected as 4 $k\Omega$ and 1nF, respectively.

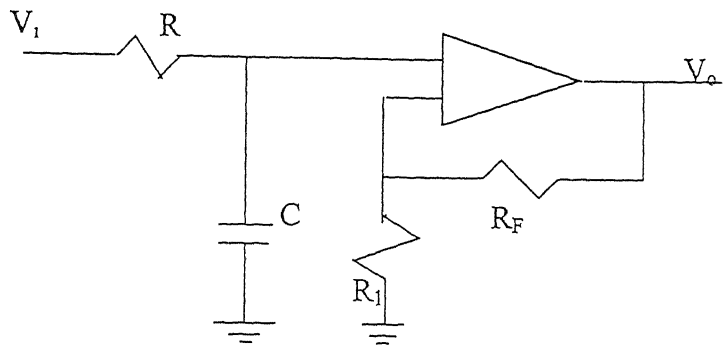


Fig.3.2: Low pass filter with 400 kHz cut off frequency

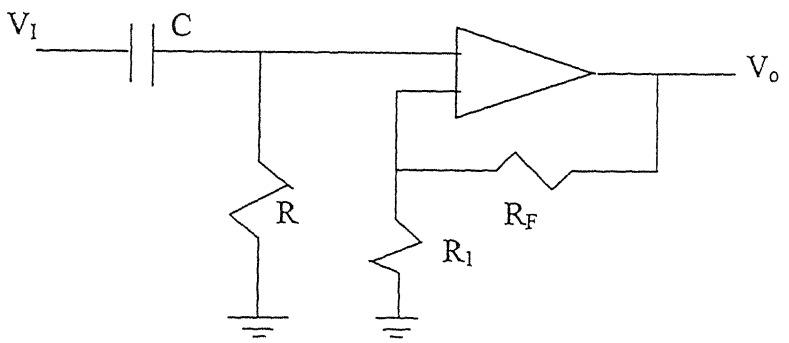


Fig.3.3: High pass filter with 40 kHz cut off frequency

Fig. 3.4 shows the effect of using the filter and the amplifier on the measured partial discharge pulses. In Fig. 3.4, channel 1 (CH1) shows the input pulses of the filter and amplifier while channel 2 (CH2) shows the output pulses. The effect of the filter and amplifier can be noticed by comparing the scaling of the two channels (channel 1was adjusted to *0.2V/cm* while channel 2 was adjusted to *5V/cm*).

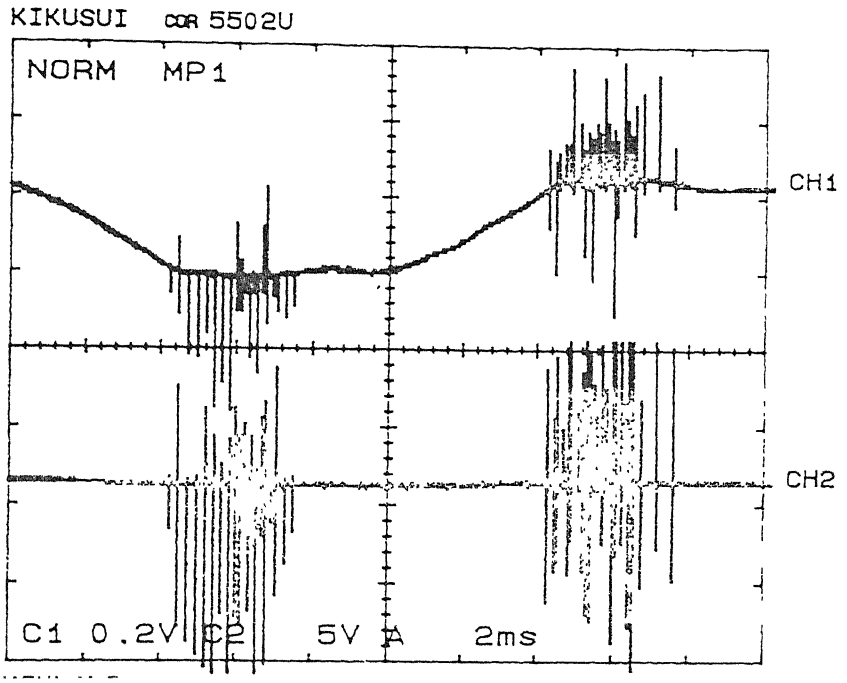


Fig. 3.4: The effect of the filter and the amplifier on the measurements
(CH1: The partial discharge pulses before the filter & the amplifier
CH2: The partial discharge pulses after the filter & the amplifier)

3.2.4 Digital oscilloscope

Kikusui COR5502U oscilloscope was used for digitizing and storing the PD data. The oscilloscope has two saving memory units, each having a 4k word capacity. The maximum sampling rate of the oscilloscope is 100 MS/s (10 ns). The signals from both the channels can be digitized whereas each channel has its own A/D converter. The input analog signal is converted into digital form by A/D converter and its output is copied onto the acquisition RAM. Then the data is transferred from the acquisition RAM to the display RAM. The oscilloscope displays it on the CRT screen with the help of D/A converter. After displaying the waveform for about one second, the next data acquisition cycle starts.

The IFO2-COR interface which is based on the standard RS-232C link, was attached onto the oscilloscope. RS-232C is used for communication between oscilloscope and peripheral terminals at relatively slow transmission. Alphanumeric communication is most frequently done using 7-bit ASCII code. Generally ASCII can be transmitted in either parallel 8-bit group (8 separate wires) or as a serial string of 8 bits, one after the other, over a single line. RS-232C was used for serial transmission. Communication between the computer and the oscilloscope was done by using full handshaking. Since the oscilloscope was working in the single triggering mode, the computer sends triggering command to the oscilloscope and then wait for a massage from oscilloscope that the waveform has been captured and digitized and it is ready for transfer. After transferring the digitized waveform to the computer, another triggering command is given for the next waveform. In the measurement set up, the waveform was the partial discharge pulses super imposed on a power frequency cycle. The number of required cycles for measurement could be controlled by the computer.

Since the partial discharge pulses occur at the maximum rising or maximum decreasing parts of the applied high voltage only and the patterns of the pulses at these parts are different, it is required to study each part individually. Therefore, a *RC* circuit was used to shift the triggering signal by 90 degree to locate the partial discharge pulses around the peak of the triggering sine wave. Figs. 3.5a and 3.5b show the position of the partial discharge pulses with respect to the applied voltage before and after using the phase shifter, respectively.

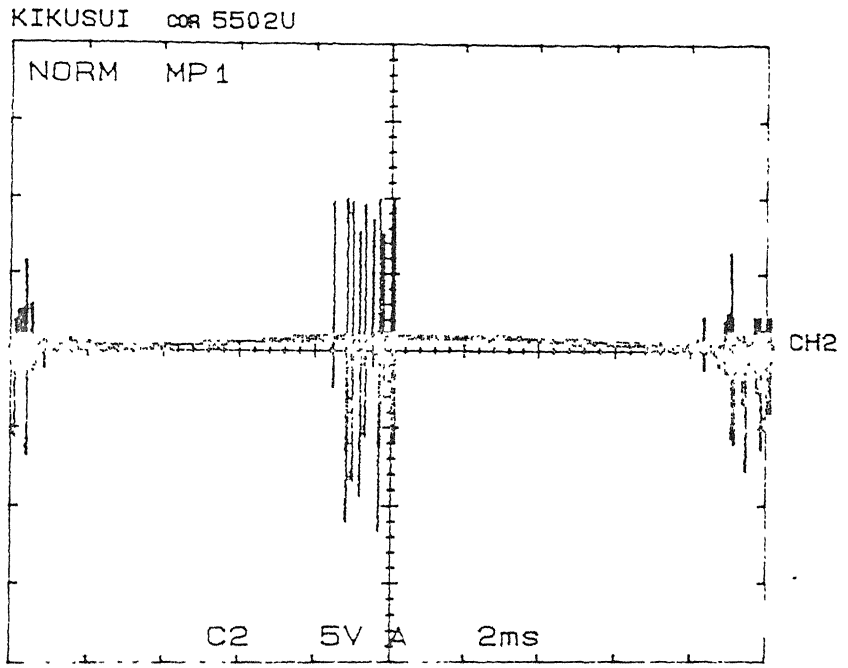


Fig. 3.5a The partial discharge pulses before the phase shifter

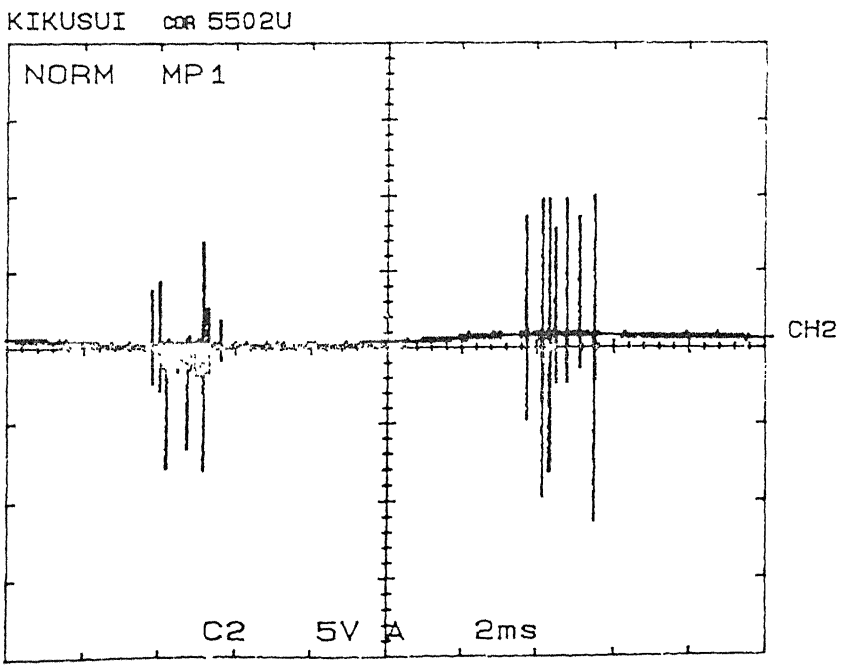


Fig. 3.5b The partial discharge pulses after the phase shifter

Fig. 3.5: The effect of the phase shifter on the partial discharge measurements

3.2.5 Personal computer

Due to the increasing trend of automatic partial discharge measurements in recent years, the use of digital system has become very popular. A personal computer offers the opportunity to store the discharge pulse sequence and to process these in the course of time or as a function of the power frequency cycle. In the present set-up, Pentium computer having 24 MB Ram, 1.6 GB hard desk and 150 MHz speed was used. For the pulses of each half cycle transferred to the personal computer, the maximum value of partial discharge pulses has been detected. All the pulses which had magnitude of less than 10 % of the maximum pulse magnitude, in the same half cycle, were ignored. Therefore, generally each pulse has been represented by only its first peak or both the first and second peaks depending upon the relative magnitude of the first peak and subsequent peaks.

3.3 Samples

In order to generate different partial discharge patterns, 6 types of standard defects were simulated with the help of simple physical models as shown in Fig. 3.6.

3.3.1 Glow corona

This was produced using point and plane electrode system with a gap distance of 15 cm in atmospheric air. The point electrode was 15 cm long and having 1.2 cm diameter. The tip was 10 cm long cone shape with angle of around 7°. A plane electrode with 7.5 cm diameter was used. Partial discharge inception voltage was around 7.5 kV.

3.3.2 Streamer corona

This was produced using rod and plane electrodes with a gap of 15 cm in the atmospheric

air. The rod electrode was 13 cm long and having 2.5 cm diameter. The tip was hemisphere with radius 1.25 cm. A plane electrode of 7.5 cm diameter was used. The partial discharge inception voltage for this electrode system was measured to be around 50 kV.

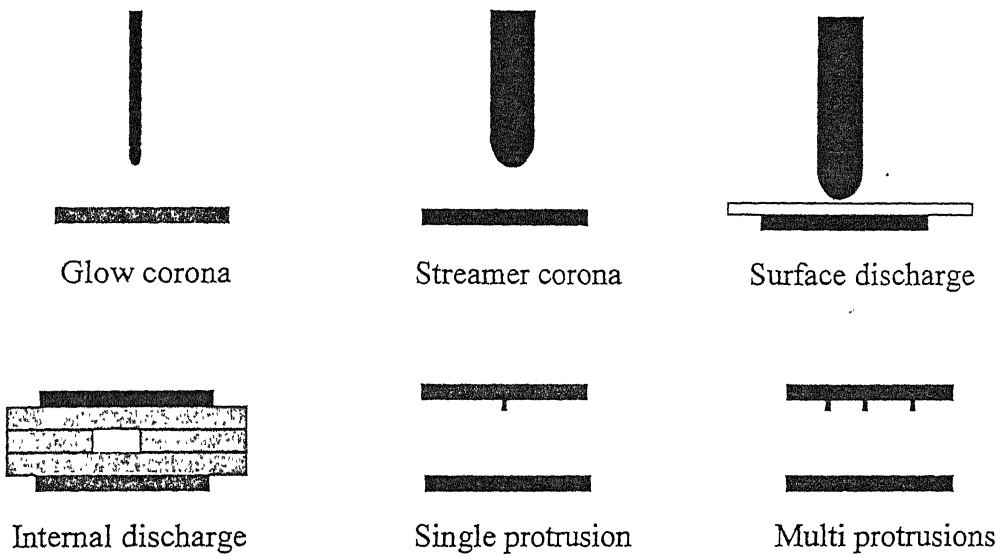


Fig. 3.6: Electrode systems for partial discharge sources created in the laboratory

3.3.3 Surface discharge in air

A glass sheet was inserted between a rod and a plane electrode system. The internal discharge was eliminated by using a thin layer of transformer oil between the glass and the grounded electrode. The inception voltage for this case was very low compared to the other electrode systems. It was about 6 kV.

3.3.4 Internal discharge

A cavity was produced by creating a hole in a bakelite sheet of thickness 1.6 mm. Two 15×15 cm sheets with centric hole were inserted between another two sheets of the same dimensions. The diameter of the cavity and its height were 4mm and 3.2mm, respectively.

Surface discharge between the bakelite sheets as well as between the sheets and the electrodes were prevented by dipping the test sample and the electrodes in transformer oil. The oil was not allowed to fill up the cavity. The inception voltage for this cavity was measured to be 28 kV.

3.3.5 Single protrusion

A small needle was used in electrode system having nearly uniform electric field to generate partial discharge in air. Two rounded edge circular plates of 12 cm diameter and with 10 cm air gap length were used. Partial discharge inception voltage was measured to be 25 kV.

3.3.6 Multi protrusions

A number of small needles were used in electrode system having nearly uniform electric field to generate partial discharge in air. The same electrode system which was used in the single protrusion case was used also in this case. Partial discharge inception voltage was measured to be around 20 kV.

Each one of these test electrode systems was subjected to voltage, raised slowly, up to 20% greater than the PD inception voltage. The latter being defined as the value of applied voltage at which the discharge once initiated do not disappear with time (self- extinction). All the plane electrodes used had a curved profile.

3.4 Conclusion

In this chapter, the partial discharge measuring circuit as well as the different electrode system used to generate the partial discharge sources have been described. Six different

partial discharge sources were generated in the laboratory. These sources include glow corona, streamer corona, surface discharge, internal discharge, single protrusion and multi protrusions. From each one of these sources, several partial discharge patterns were generated. In the subsequent chapters, the textures analysis algorithms will be used to examine their ability to distinguish between the different partial discharge sources created in the laboratory.

Chapter 4

Classification of Partial Discharge Patterns Using Minimum Distance Classifier

4.1 Introduction

Chapter 2 gave a review for texture analysis and revealed that texture information, which has been used successfully in various applications, has potential for classification of partial discharge sources. However, the main questions to be addressed are:

(i) How effective the texture information can be used for partial discharge sources classification?

(ii) What is the effect of changing the number of cycles on the classification accuracy of partial discharge sources?

(iii) Which features have good discriminating power?

(iv) The minimum number of features which can be used at a time to achieve a certain desired classification accuracy?

This chapter evaluates the ability of four texture analysis algorithms for classifying partial discharge sources. The algorithms which have been examined are the spatial gray level dependence method (SGLDM), the gray level difference histogram method (GLDHM), the

gray level run length method (GLRLM) and the power spectral method (PSM). Only two directions have been used in this thesis; the horizontal and the vertical directions. These two directions have been selected in the sense that one can investigate the relationship between the adjacent pulses in the same cycle (horizontal direction) as well as between the pulses in the same phase window (vertical direction).

A comparison between the features based on these algorithms and some statistical features calculated from the conventional phase resolved patterns analysis method (q-φ-n distribution) has been carried out. The optimal features of each of these algorithms have been selected according to their classification accuracy. Experimental tests were carried on samples with different discharge sources to establish the discrimination power of these algorithms using minimum distance classifier [Tou and Gonzalez, 1974].

4.2 Selection of the Partial Discharge Patterns

In order to apply the texture analysis algorithms, the proper image of each defect has to be created. This requires a selection of proper sampling technique and the sample size as described below.

4.2.1 Sampling techniques

The first step in the generation of partial discharge patterns is the selection of only few power frequency cycles out of a series of cycles generated through the experiment. In other words, sample for each partial discharge source has to be selected. In literature, three sampling techniques are available [Bhattacharyya, 1977]:

(i) Simple random sampling

(ii) Stratified random sampling

(iii) Systematic sampling

In the first technique, one could measure randomly N cycles to construct a partial discharge image. In the second technique also, the concept of simple random sampling technique is used. However, to improve the accuracy to be attained for different types of population, the complete data is divided into set of homogenous segments (sub-population) and then independent simple random samples are drawn from the individual sub-population. In the third technique, the purpose is to spread the sample over the population. This technique divides the population into several sub-population and then randomly selects a cycle from the first sub-population (say cycle number K) and the cycles having the same number in the other sub-populations. Thus, to use the stratified random sampling and systematic sampling techniques, the size of population should be known. In the case of partial discharge patterns, the size of population (number of cycles which can be measured) is infinity. So, simple random sampling technique is considered to be the most suitable choice for this case. In the experiment, N cycles were selected and measured randomly. However, it involved a constant time lag equal to the digitising time and the time of transferring the data to personal computer.

If the partial discharge measurements are taken through N cycles of the applied power frequency high voltage waveform and each cycle is divided into M windows, an image of dimension $N \times M$ pixels is obtained. Replacing the gray level values used for images interpretation by the magnitude of the partial discharge pulses, the texture analysis

algorithms can be applied to extract discriminating features for different partial discharge sources. In other words, the magnitude of each pulse can be compared with the magnitude of the adjacent pulses (right and left) in the same cycle as well as with the magnitude of pulses in the same window (phase angle position) in the other cycles (up and down). Thus only two directions are required, the horizontal (in the same cycle) and the vertical (different cycles). In the present work, total 4096 windows were selected in each cycle and the number of cycles (N) were varied between 2 and 32 for each partial discharge source. Since the behaviour of partial discharge is different in the positive and negative half cycles, the texture features were calculated for each half cycle individually.

4.2.2 Training data refinement

For each technique, all features have been calculated for a sample size of 30 patterns belonging to each class or partial discharge source by using number of cycles varying between 2 and 32. The mean values and the standard deviations of all the features have been calculated for each technique individually. Sometimes it is possible that, in the training data, some aberrant patterns are included. Automatic methods to remove such patterns have been suggested (Buttner et al., 1989). In the present investigation, this refinement to the training patterns was performed by removing those patterns whose distance to the actual mean was greater than a predefined limit. In fact, such patterns correspond to a low probability of belonging to a class defined by its mean and covariance matrix. Thus an i^{th} pattern characterised by feature j is removed from the training data if the following inequality is fulfilled.

$$\left|X_{ij} - \mu_j\right| > K\sigma_j \quad . \tag{4.1}$$

where

μ_j is the mean value of feature j

σ_j is the standard deviation of feature j

K is specified parameter for cut-off limit

In the present study K was taken as 3. Table 4.1 shows a typical results for GLDHM features calculated for two cycles. Similar tables were generated for each technique and the five cases corresponding to 2, 4, 8, 16, 32 cycles.

Table 4.1 Mean and standard deviation of GLDHM features

Features	with out refinement			with refinement		
	Mean value	SD	no. of patterns	Mean value	SD	No. of patterns
1	28.23897	3.558595	30	27.70894	2.094585	29
2	93.63261	2.275512	30	93.63261	2.275512	30
3	23.02265	1.878695	30	23.02265	1.878695	30
4	35.45954	8.339054	30	34.06636	3.422692	29
5	96.77002	1.177092	30	96.77002	1.177092	30
6	89.91356	12.8845	30	92.03762	5.635258	29
7	30.0577	9.975033	30	28.27342	2.0333	29
8	91.88288	7.293665	30	93.20567	0.853631	29
9	25.49458	2.700341	30	25.13637	1.8882	29
10	30.37757	13.38121	30	27.97679	2.522866	29
11	95.70605	4.711396	30	95.70605	4.711396	30
12	2.450556	12.37675	30			
13	99.55195	2.288191	30	99.55195	2.288191	30
14	0.194182	0.948212	30			
15	3.701664	18.23956	30	-2.15E-05	1.59E-08	19
16	99.77503	1.15135	30	99.77503	1.15135	30
17	-71.9008	47.4012	30	-100	6.37E-08	22
18	3.420175	18.24341	30			
19	98.54305	7.806394	30	99.9681	0.130966	29
20	0.355971	1.863411	30			
21	3.509818	18.23434	30	-1.02E-05	1.33E-09	19
22	99.09216	4.887973	30	99.09216	4.887973	30

4.2.3 Sample size determination

The vector of features calculated for each partial discharge source using N cycles represents a pattern in the feature space. Each pattern class consists of several patterns to represent that class. However, an adequate number of patterns have to be used in each pattern class. In determining the sample size (i.e. the number of patterns) in a statistical experiment, the following questions should be addressed [Pfaffenberger and Patterson ,1977]:

(i) How close the estimate is required to be to the true value of the population parameter?

(ii) How certain is it desired that the estimate will be within the selected number of units of the value of the parameter?

For example, it may be specified that the patterns estimate of the population mean μ_1 should not be more than ten units away from the true value of mean μ with a confidence of 95 percent. This implies that one is willing to take a 5 percent chance that the sample or the partial discharge patterns measured of size n will produce an estimated mean μ_1 which will be more than ten units away from the actual value μ. The confidence coefficient depends on how many times of standard deviation are included. The following table shows the relationship between the confidence coefficient and the standard deviation σ.

Confidence Coefficient (Z)	0.90	0.95	0.99
Times of Standard Deviation (σ)	1.64	1.96	2.58

For example, a confidence of 95 percent could be mathematically expressed as:

$$prob(\mu_1 - 1.96\sigma \le \mu \le \mu_1 + 1.96\sigma) = 0.95 \qquad (4.2)$$

Generally, the standard deviation of the population is unknown. Therefore it is replaced by $s / n^{0.5}$ where s is the sample standard deviation and n is the sample size. For an error, which should not exceed a value (d), one can get the following equation:

$$Z\frac{s}{\sqrt{n}} = d \qquad (4.3)$$

Therefore,

$$n = \left[\frac{Z\sigma}{d}\right]^2 \qquad (4.4)$$

This equation determines the required sample size i.e. the number of partial discharge patterns which will represent one defect or partial discharge source for a confidence coefficient (Z) and the absolute error (μ - μ_1) not exceeding d. In deciding whether to compute for 95% confidence interval or for 99% confidence interval, it may seem ridiculous to settle for a lower level of confidence when higher level could be obtained. However, the old adage "*you must give up something to get something*" remain true here because a 95% interval extends 1.96σ to either side of the mean μ_1. Similarly for 99% confidence the interval will be 2.58σ to either sides of mean μ_1. This means that the higher the desired degree of confidence, the longer the resulting interval. If the length of the interval is specified for measuring precision or accuracy, then the confidence level of the interval is inversely related to its precision for the same sample size [Devore, 1987]. Otherwise, to achieve the same precision the sample size should be increased. For the features of the texture analysis algorithms, the variance is not the same. Therefore, it is important to select the sample size according to the feature which has maximum coefficient

of variation α. The coefficient of variation of i^{th} feature is its standard deviation divided by its mean value as given below:

$$\alpha(i) = \frac{\sigma(i)}{mean(i)} \tag{4.5}$$

The final size of the samples in each technique has been calculated according to the feature which has the maximum coefficient of variation. For computing the coefficient of variation for each feature, the mean and standard deviation values obtained after refinement has been used. The sample size has been calculated for confidence intervals 1.64 (90%) and 1.96 (95%) as well as for 5% and 10% errors from the means. Tables 4.2 and 4.3 show the estimated sample size (number of patterns in each pattern class) for the different techniques (GLDHM, SGLDM, GLRLM, PSM, q-φ-n).

Table 4.2 Sample size for 1.64 confidence interval.

Techniques	Number of cycles and acceptable error from the mean value									
	2 cycles		4 cycles		8 cycles		16 cycles		32 cycles	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
GLDHM	20	5	14	3	9	2	9	2	10	2
SGLDM	20	5	14	3	9	2	9	2	11	2
GLRLM	5	1	8	2	8	2	13	3	17	4
PSM	672	168	177	44	84	21	84	21	34	8
q-φ-n	18	4	38	9	15	3	25	6	31	7

Table 4.3 Sample size for 1.96 confidence interval.

Techniques	Number of cycles and acceptable error from the mean value									
	2 cycles		4 cycles		8 cycles		16 cycles		32 cycles	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
GLDHM	29	7	21	5	12	3	13	3	15	3
SGLDM	29	7	21	5	12	3	13	3	15	3
GLRLM	7	1	12	3	11	2	19	4	25	6
PSM	960	240	254	63	121	30	120	30	49	12
Q-φ-N	25	6	55	13	22	5	36	9	45	11

From these tables the following conclusions can be drown:

(i) A confidence interval of 1.96 and error of 5% from the mean gives a reasonable sample size for most of the techniques.

(ii) For the GLDHM, SGLDM and GLRLM techniques, the maximum sample size was less than 30 patterns. This is due to the smaller values of standard deviation of the patterns in the same pattern class.

(iii) For PSM the variation between the patterns was very high resulting in large sample size. However, this variation reduced with increasing the number of power frequency cycles. This is due to the reason that with the increased number of power frequency cycles, the information included in each pattern increases. Hence the patterns represent that pattern class very well and all the patterns become closer which reduce the coefficient of variation.

(iv) The above conclusion, however, does not seem to be applicable to the other techniques. Analysing the features having the maximum coefficient of variation for these techniques, it is found that for PSM technique, feature number 3 has maximum variation irrespective of number of cycles used. However, for GLDHM for example, the features not same and were 9, 9, 3, 9, 4 for 2, 4, 8, 16, 32 cycles respectively. These feature results in sample size of 29, 21, 12, 13, 15 patterns. But it is interesting to notice that for the same feature like number 9, the sample size has decreased as in the case PSM with increasing the number of power frequency cycles. It was 29, 21, 13 patterns for 2, 4, 16 cycles respectively. This analysis implies that, for the same feature, increasing the number of cycles increase the

discriminating information which can be extracted by that feature. Similar observation have been made for the other techniques.

From the above results, a sample size of 30 patterns were selected to represent each class which satisfies requirement of most of the techniques.

4.3 Minimum Distance Classifier

The traditional methods for classification mainly follow two approaches: unsupervised and supervised. The unsupervised approach attempts to identify a pattern in the feature space. However, it may result in groupings that have no clear meaning from the user's point of view. Having established this, the analyst then tries to associate an information class with each group. The unsupervised approach is often referred to as clustering. In the supervised approach of classification, the user supervises the pattern categorisation process by specifying, to the computer algorithm, numerical descriptors of the various partial discharge sources. To do this, representative samples of known partial discharge sources, called training patterns, are used to compile a numerical interpretation key that describe these patterns. Each pattern in the data set is then compared numerically to each category in the interpretation key and labelled with the name of the category it looks most like. In the unsupervised approach the user defines useful information categories and then examines their separability whereas in the supervised approach one first determines separable classes and then defines their information. In the present work, to identify the source of particular partial discharge, a supervised classification technique has been used to label an unknown source by using trained classifier.

Classification has been done according to the similarity between the pattern and the pattern populations. The distance between any pattern and the pattern populations could be used as a measure of similarity in a sense that the smaller the distance, the greater the similarity. Classification has been done by using minimum distance (Mahalanobis distance) classifier [Tou and Gonzalez, 1974]. The major difference between the Euclidean distance and Mahalanobis distance lies in the use, by the latter, of the sample covariance information. In other words, Euclidean distance measure creates a circle of points equidistant from the mean value, while the Mahalanobis distance measure creates an elliptical shaped region about the mean value. The ellipse is elongated to account for correlation between variables. If one feature has a larger variance than the other, it receives less weight. The Mahalanobis distance is given by:

$$D = (x - m)^T C^{-1} (x - m) \qquad (4.6)$$

where

C is the covariance matrix of a pattern population

m is the mean vector

x represents an unknown pattern.

The minimum distance classifier is a parametric classifier which relies on specifying the probability distribution of the class values by representative function. In this case, the function is Gaussian and the mean and covariance are its parameters. The mean vector in multidimensional space locates the single peak of the probability distribution and the covariance specifies the region of influence of the probability distribution in each dimension. Since the normal distribution provides a good approximation to many naturally

occurring phenomena, it is expected that the partial discharge classes will also be normally distributed. The histograms of all classes by using different features of all the techniques is shown in Appendix -1.

Overall classification accuracy has been calculated from the confusion matrix which contains information about the correct classification and misclassification of all classes. Confusion matrix is an $A \times A$ matrix, where A is the number of classes or partial discharge sources. The rows in the matrix represent the assumed true classes, while the columns are associated with the classified partial discharge sources. In a confusion matrix, the diagonal elements represent the observations which agree with the true classes and the non-diagonal entries represent cases where observations do not agree. The degree of agreement is given by:

$$p = \frac{(X_{11} + X_{22} + ... + X_{AA})}{\sum_{i=1}^{A} \sum_{j=1}^{A} X_{ij}} \qquad (4.7)$$

$$p = \frac{(X_{11} + X_{22} + ... + X_{AA})}{N} \qquad (4.8)$$

where, N is the total number of elements in the matrix, p is proportion of the patterns that agree, and X_{ii} is the diagonal elements of the $A \times A$ confusion matrix. For 100 percent correct classification this matrix should be diagonal.

4.4 Partial Discharge Sources Classification Results

The main aims of the study presented in this chapter have been:

(i) To investigate the discriminating power of the texture analysis algorithms.

(ii) To identify the features which have the best discriminating power.

(iii) To determine the minimum number of features which can be used at a time to give an acceptable classification accuracy.

(iv) To investigate the effect of changing the number of cycles used to construct the partial discharge patterns on the classification accuracy.

To achieve the first three aims, the classification of PD sources has been done in three stages, the first by using the full set of features, the second by using each feature in sequence to determine the optimum features, and the last stage by using combinations of two features to select the best combinations. Achieving the third aim was relatively difficult due to the correlation between the features. One has to investigate all the possible feature combinations. Investigations were carried out to determine the classification accuracy of each method as well as the effect of changing the number of cycles in each pattern on the classification accuracy. For the cases of 2, 4, 8, 16 and 32 cycles, 30 partial discharge patterns were used for training and testing the minimum distance classifier. As shown in Fig. 4. 1a, it was found that:

(i) All the techniques achieved 100 % classification accuracy for different number of cycles.

(ii) PSM has 96.66% classification accuracy for the case of 2 cycles and it increased to 100% for the higher number of cycles.

(iii) Increasing the number of cycles has no effect on the classification accuracy of the different techniques.

(ii) To identify the features which have the best discriminating power.

(iii) To determine the minimum number of features which can be used at a time to give an acceptable classification accuracy.

(iv) To investigate the effect of changing the number of cycles used to construct the partial discharge patterns on the classification accuracy.

To achieve the first three aims, the classification of PD sources has been done in three stages, the first by using the full set of features, the second by using each feature in sequence to determine the optimum features, and the last stage by using combinations of two features to select the best combinations. Achieving the third aim was relatively difficult due to the correlation between the features. One has to investigate all the possible feature combinations. Investigations were carried out to determine the classification accuracy of each method as well as the effect of changing the number of cycles in each pattern on the classification accuracy. For the cases of 2, 4, 8, 16 and 32 cycles, 30 partial discharge patterns were used for training and testing the minimum distance classifier. As shown in Fig. 4. 1a, it was found that:

(i) All the techniques achieved 100 % classification accuracy for different number of cycles.

(ii) PSM has 96.66% classification accuracy for the case of 2 cycles and it increased to 100% for the higher number of cycles.

(iii) Increasing the number of cycles has no effect on the classification accuracy of the different techniques.

In fact, by changing the number of cycles, the texture features' values have changed. However, the pattern classes still separate and thus resulting in high classification accuracy. To investigate the ability of features created from the texture analysis algorithms to generalise the extracted information from the partial discharge patterns, different patterns were used for training and testing of the minimum distance classifier. In other words, patterns created from different number of cycles were used for testing and training. For the different techniques, the patterns created using 2, 4, 8, 16 and 32 cycles, 30 patterns each, were used for the minimum distance classifier training, whereas the testing has been done by the 30 patterns created from the case of 32 cycles only. The classification accuracy for this case is shown in Fig. 4. 1b. From this figure, it can be observed that:

(i) SGLDM and GLDHM are having almost the same classification accuracy for the different number of cycles used for training.

(ii) With increasing the difference between the number of cycles used for training and testing, the classification accuracy of the different techniques is reduced.

(iii) Reducing the number of cycles used for training has small effect on the classification accuracy of the SGLDM and GLDHM. Whereas, GLRLM and PSM are more affected by reducing the number of cycles.

(iv) The features created from the phase resolved pulse height analysis (q-ϕ-n distributions) are highly affected by increasing the difference between the cycles used for training and testing.

It is clear from Fig. 4.1b that the texture analysis algorithms are able to extract the discriminating information between the different partial discharge sources. This information slightly depends upon the number of cycles used to construct the partial discharge patterns. However, this property is not valid for the conventional methods (q-ϕ-n distributions).

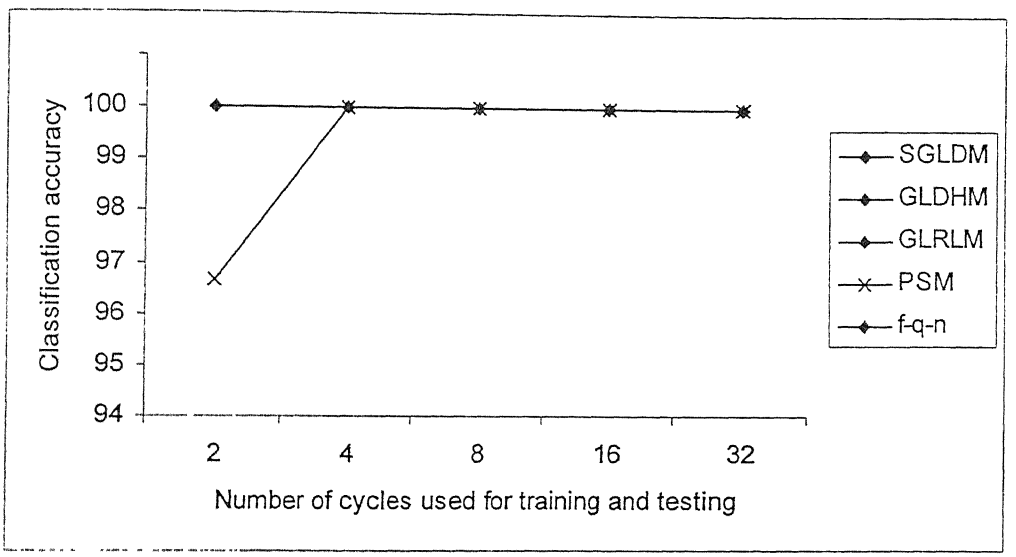


Fig. 4.1a: Classification accuracy of different techniques using the same patterns for training and testing

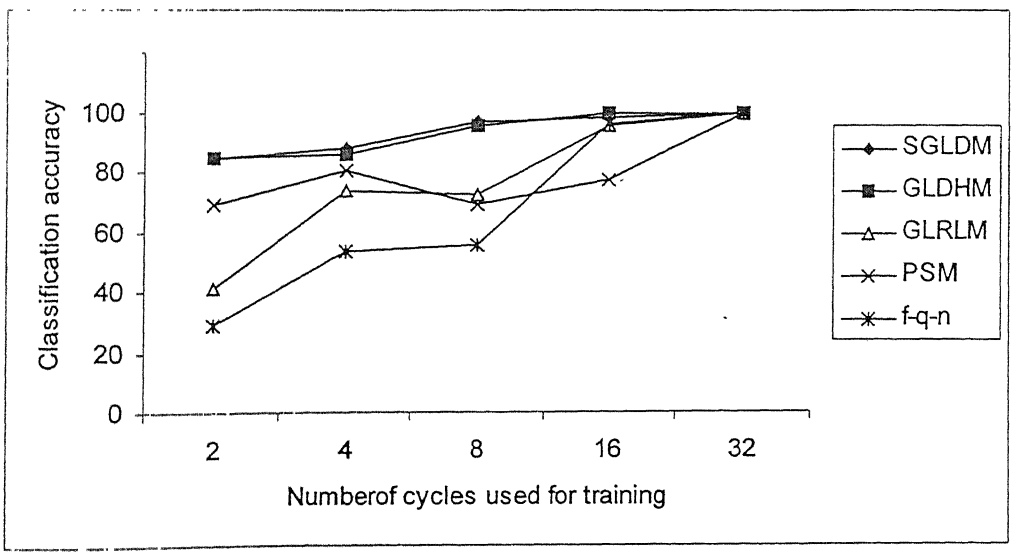


Fig. 4.1b: Classification accuracy of different techniques using different patterns for training and testing

The classification accuracy shown in Fig. 4.1a and 4.1b were achieved by using all the features in each technique. The large number of features, however, make the classification process quite time consuming. Therefore, the features which have little contribution to the classification process can be discarded. The process of features selection can be done according to the relative classification accuracy of the features. This has been established for the four texture analysis algorithms in addition to the ϕ-q-n distributions. For each pattern class, 15 patterns were used for training and 30 patterns (including the training patterns) were used for testing. All patterns were created by using the same number of cycles for training and testing.

4.4.1 Spatial gray level dependence method (SGLDM)

As mentioned in chapter 2, SGLDM has seven features in each direction. These features are mean, energy, contrast, entropy, inverse difference moment, variance and the correlation. The scanning of each pattern horizontally and vertically for the positive and negative half cycle individually results in total of 28 features. Features 1-7 and 8-14 correspond to the horizontal and the vertical directions for the positive half cycle measurements. Whereas, features 15-21 and 22-28 correspond to the horizontal and the vertical directions of the negative half cycle measurements. Each one of the 28 features were considered individually for training and testing. Fig. 4.2 shows the classification accuracy of all the features using the six partial discharge sources discussed in chapter 3. It can be observed from Fig. 4.2 that:

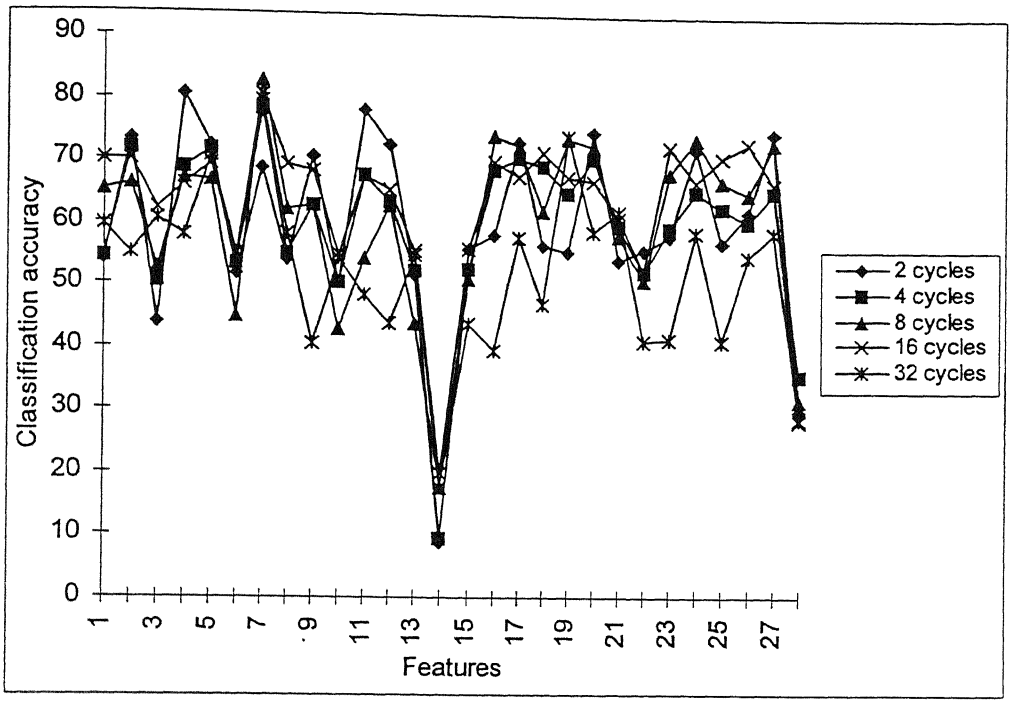


Fig. 4.2: Classification accuracy of SGLDM features

(i) The relative classification accuracy of the features depend upon the number of cycles used to construct the patterns.

(ii) Only the feature number 7 (correlation) is almost independent of the number of cycles. It has the maximum classification accuracy in the case of 4, 8, 16, and 32 cycles varying between 78% to 82%.

(iii) The discriminative power of the correlation between the pulses is more clear in the horizontal direction of the positive half cycle (features 7).

(iv) The discriminative power of the correlation between the pulses in the vertical direction for the positive and negative half cycles is very poor. It varies between 8% to 20% with changing the number of cycles for the positive half cycle (features 14). For the negative half cycle, it varies

between 28% to 35% (feature 28).

(v) Classification accuracy for the mean (features 1, 8, 15, 22) and the variance (features 6, 13, 20, 27) are the same for the horizontal and vertical directions in each half cycle individually.

(vi) The classification accuracy due to any feature in the horizontal direction is greater or at least equal to its classification accuracy in the vertical direction in both the half cycles.

4.4.2 Gray level difference histogram method (GLDHM)

In case of GLDHM, as given in chapter 2, there are 5 features in each direction for each half cycle. These features are mean, energy, contrast, entropy and inverse difference moment. For the horizontal and vertical directions of the positive and negative half cycles, one gets 20 features. To consider the pulses polarity, another feature was introduced called as the polarity factor. This additional feature is the difference between the average of positive pulses and the average of negative pulses divided by the addition of the average of positive pulses and the average of negative pulses as given below:

$$Polarity\,factor = \frac{\frac{\sum q^+}{n^+} - \frac{\sum q^-}{n^-}}{\frac{\sum q^+}{n^+} + \frac{\sum q^-}{n^-}} \qquad (4.9)$$

where n^+ and n^- are the number of positive and negative pulses. The value of this factor varies between –1 and 1. It will be positive if the average of the positive pulses is greater

than the average of the negative pulses and vice versa. This feature is the same for horizontal and vertical directions. Hence, it was considered in the horizontal direction only which makes the total features as 22. Fig. 4.3 shows the classification accuracy of GLDHM features when considered individually. Features 1-6 were calculated for the horizontal direction of positive half cycle and features 7-11 were calculated for the vertical direction of the same half cycle. Features 12-17 and 18-22 were calculated for the negative half cycle in horizontal and vertical directions, respectively. From this figure, it is clear that:

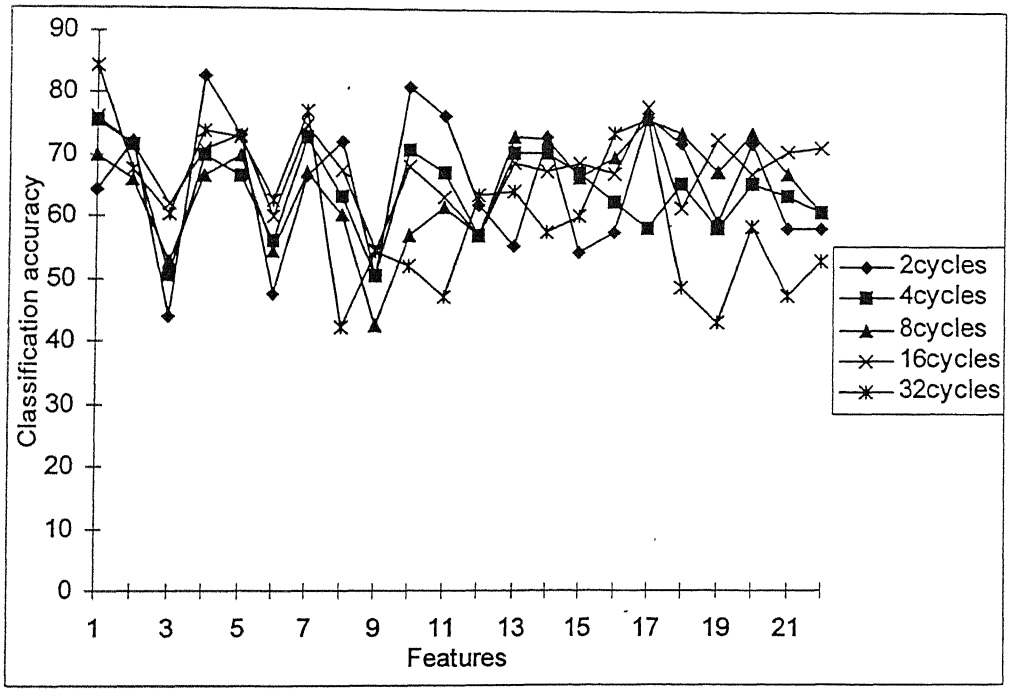


Fig. 4.3: Classification accuracy of GLDHM features

(i) The relative classification accuracy of the features depend upon the number of cycles used to construct the patterns.

(ii) In the positive half cycle, the classification accuracy of the features used in the horizontal direction is higher than the classification accuracy of the same

features when used in the vertical direction, except the case of 2 cycles.

(iii) In the negative half cycle, there is no clear relationship between the classification accuracy of the features and the direction.

4.4.3 Gray level run length method (GLRLM)

In case of the GLRLM, there are five textural features as mentioned in chapter-2 which resulted in total of 20 features when calculated for the horizontal and vertical directions in the positive and negative half cycles. These features are short run emphasis, long run length emphasis, gray level non-uniformity, run length non-uniformity and run percentage. As shown in Fig. 4.4, features 1-5 correspond to the horizontal direction of positive half cycle and features 6-10 correspond to the vertical direction of the same half cycle. Features 11-15 and 16-20 correspond to the negative half cycle in horizontal and vertical directions, respectively. From this figure, the following can be observed:

(i) The relative classification accuracy of the features depend upon the number of cycles used to construct the patterns.

(ii) The short run emphasis is equally or more effective in classification when used in the horizontal direction for both the positive and negative half cycles (features 1, 6) in comparison with the same feature when used in vertical direction of the positive and negative half cycles (features 6, 16) except for the case of 2 cycles. On the other hand the same feature is, in general, more effective in the positive half cycle compared to the negative one.

(iii) The long run emphasis is equally or more effective in classification when

used in the vertical direction for positive and negative half cycles (features 7, 17) in comparison with the same feature when used in the horizontal direction of the positive and negative half cycles (features 2, 12) except for the case of 32 cycles. On the other hand, the same feature is generally more effective in the negative half cycle compared to the positive one.

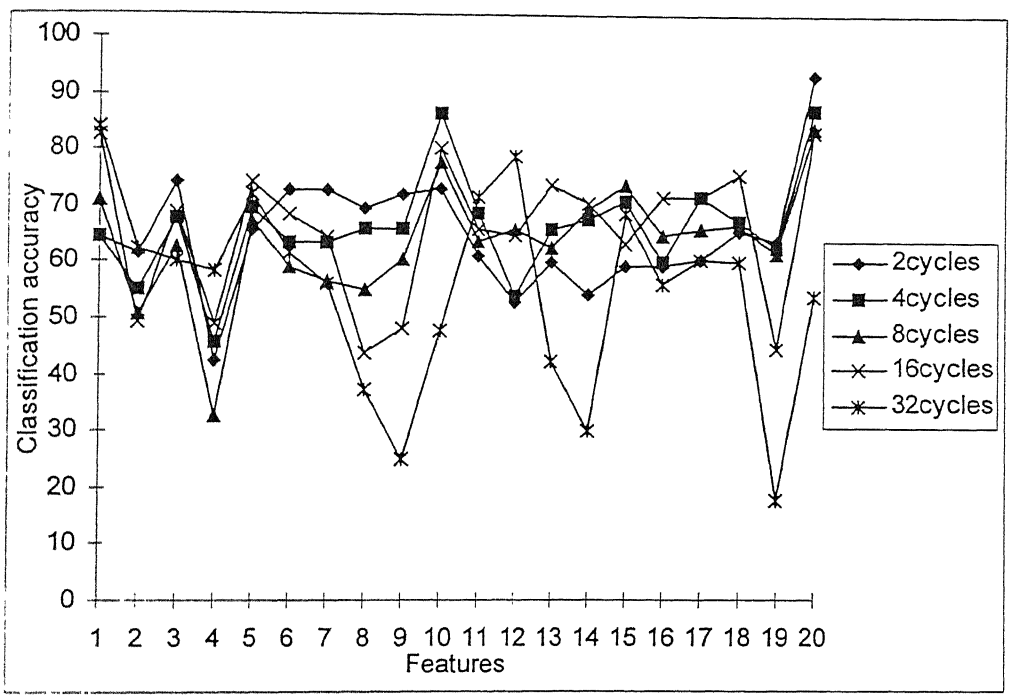


Fig. 4.4: Classification accuracy of GLRLM features

(iv) The gray level non-uniformity is more effective in classification when used for the horizontal direction of the positive half cycle (feature 3) or the vertical direction of the negative half cycle (feature 18).

(v) The run length non-uniformity is more effective in classification when used for the vertical direction of the positive half cycle (feature 9) except for the case of 32 cycles. It is also effective in the horizontal direction of the

negative half cycle (feature 14) except for the case of 2 cycles.

(vi) The classification accuracy of the run percentage is higher if it used in the vertical direction of the positive and negative half cycles (features 10, 20) except for the case of 32 cycles.

4.4.4 Power spectrum method (PSM)

Since the transformed image will have the same size as the original one, in case of PSM the number of features is not constant. It depends upon the number of cycles used to construct the partial discharge patterns. For example, Fig. 4.5 shows the power spectrum of a glow corona. Actually this figure was calculated for 128 cycles with 2048 phase windows for each cycle. The transformed image was adjusted to bring the average value at its centre, so that the image is symmetrical around its centre. Since the variation being only around the centre of the image, for clarity only 32 cycles by 32 phase windows from the centre of the full transformed image are shown in Fig. 4.5. The features generated from the transformed image are the average values of number of slits (say H) in the horizontal or vertical directions. Since there are 2048 phase windows per cycle, the number of features in the horizontal direction are quite large and depends on the step size H. In this study, it was selected to be fixed and equal to 6 for all the number of cycles. The number of features in the vertical direction were varying according to the number of cycles. Therefore the number of features considered were 14, 14, 16, 18 and 24 for 2, 4, 8, 16 and 32 cycles, respectively. Half of these features were calculated from the transformed image for the positive half cycle while the other half from the negative half cycle. In each half of these features, the last six are those which were calculated for the horizontal direction. The

classification of these features are shown in Fig. 4.6, which provides the following observations:

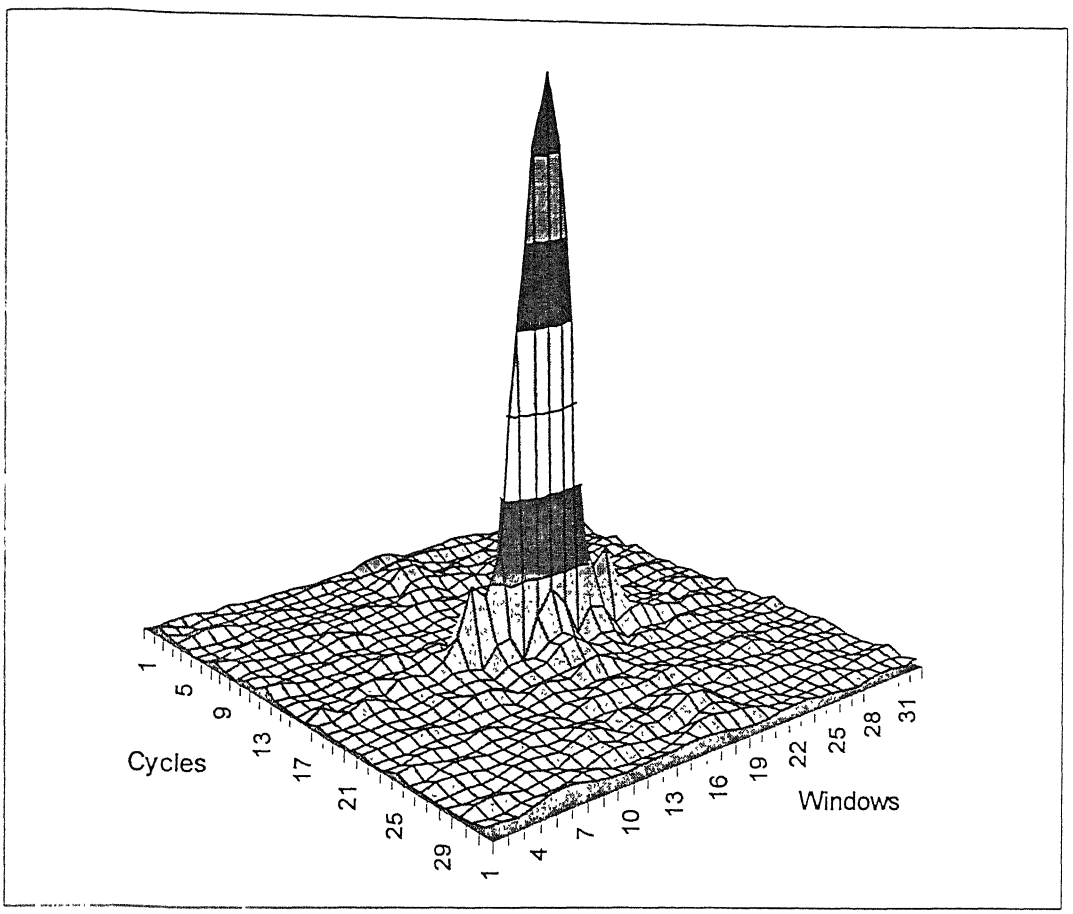


Fig. 4.5: Fourier transform of Glow corona

(i) The relative classification accuracy of the features depend upon the number of cycles used to construct the patterns.

(ii) The features which are calculated from the horizontal slits (features 1, 1, 1-2, 1-3, 1-6 for 2, 4, 8, 16, 32 cycles, respectively) are generally more effective in the classification especially those which are nearer to the image centre.

(iii) The features which are calculated for the central slits in the vertical

direction (features 2, 9 for 2, 4 cycles, features 3, 11 for 8 cycles, features 4, 13 for 16 cycles and features 7, 19 for 32 cycles) has the lowest classification accuracy.

(iv) The classification accuracy of any feature in the negative half cycle (the second half of features) is greater than the corresponding feature in the positive half cycle (the first half of features) except for the case of 32 cycles it is almost same.

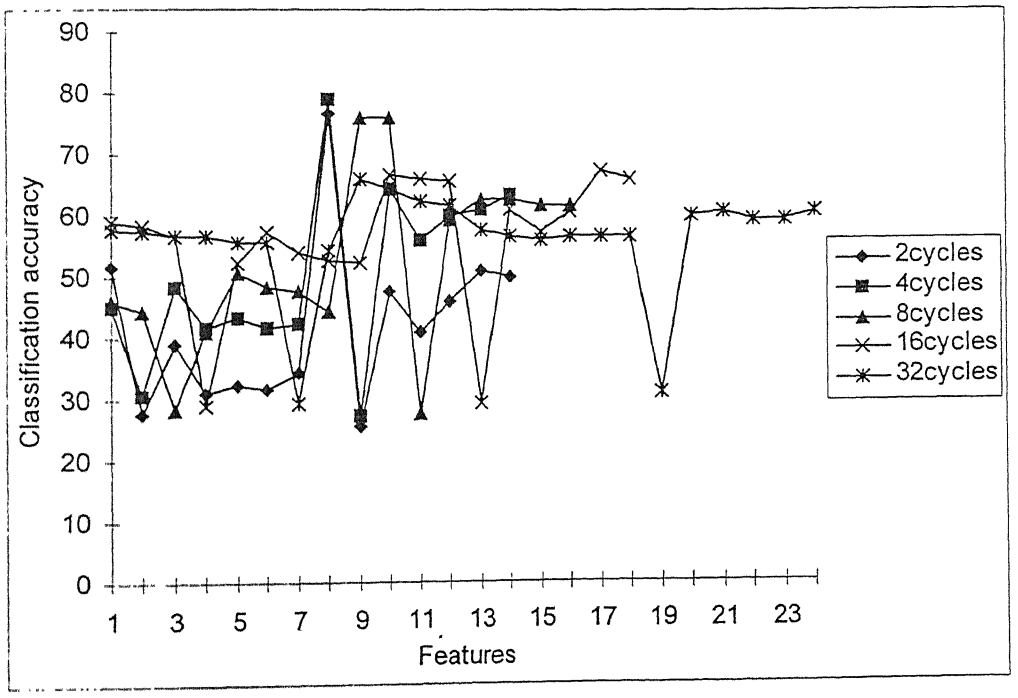


Fig. 4.6: Classification accuracy of PSM features

4.4.5 q-ϕ-n Distributions

In case of q-ϕ-n distributions, there are 7 features for each pattern as given in chapter 1. These features are the skewness, the kourtosis, the number of peaks for each half cycle in addition to the correlation between the positive and negative half cycles. These features

were calculated for three distributions and thus resulted in total of 21 features. Features 1-7 correspond to the number of pulses in each phase window (n-ϕ distribution), features 8-16 correspond to the average value of partial discharge pulses in each phase window (q_{av}-ϕ distribution) and features 15-21 correspond to the maximum pulse magnitude in each phase window (q_{max}-ϕ distribution). Fig. 4. 7 shows the classification accuracy of these features. It can be seen from this figure that:

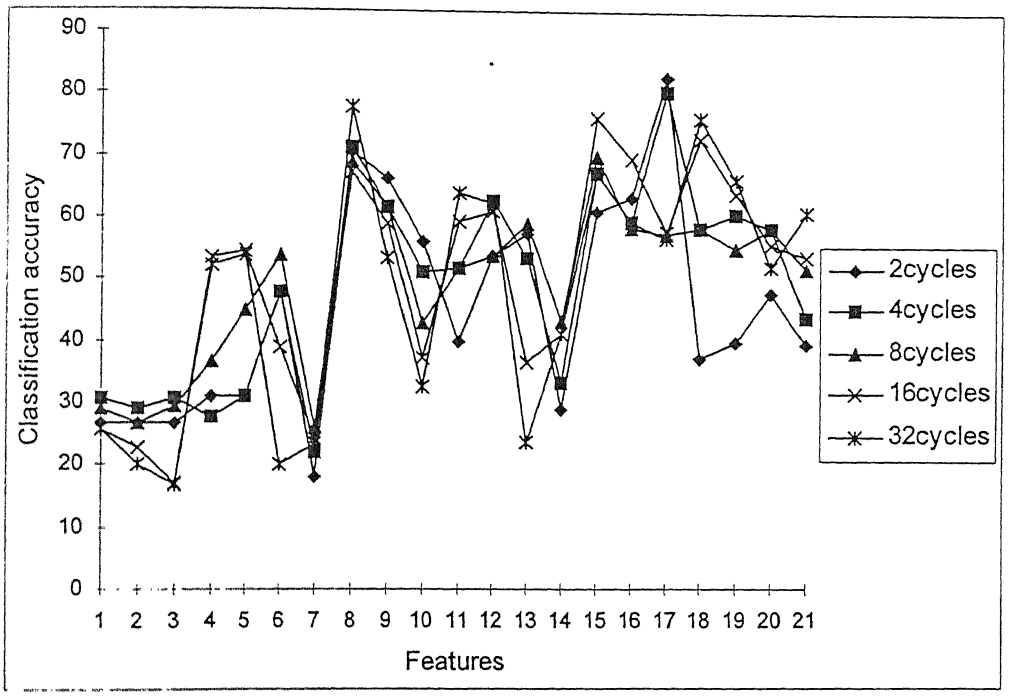


Fig. 4.7: Classification accuracy of q-φ-n features

(i) The first distribution (n-ϕ distribution) has very poor classification accuracy (features 1-7) compared to the other distributions. The maximum classification accuracy of all the features calculated from this distribution was less than 55%

(ii) The classification accuracy of the correlation between the positive and

negative pulses was very poor in the first and second distribution (features 7, 14) while it slightly improved in the third distribution (feature 21).

(iii) The third distribution (q_{max}-ϕ distribution) has the maximum classification accuracy (features 15-21).

(iv) For the positive half cycle, the skewnwss (features 1, 8, 15) is more effective in classification as compared to the kourtosis (features 2, 9, 16).

4.4.6 Effect of increasing the number of features used for classification

From the results of the classification accuracy shown above for the different features individually, it is clear that considering one feature at a time, in any of the techniques, is not sufficient to get a reasonable classification accuracy. In fact the best classification accuracy achieved with all the methods was below 85% except with the GLRLM which slightly exceeded 90%. Hence, it is important to consider more than one feature at a time to improve the classification accuracy. Because of the correlation between the different features, one can not directly select the best two features in each technique and use them together as the best combination to increase the classification accuracy. Therefore, the best combinations have been established with each technique by investigating the discriminating power for all possible combinations of two features. The results of the classification accuracy were obtained with all the texture analysis techniques for 2, 4, 8, 16 and 32 cycles. However, the classification accuracy of all the combinations using patterns constructed from only two cycles for the SGLDM, GLDHM, GLRLM, PSM, and q-ϕ-n distributions are shown in Figs. 4.8 to 4.12, respectively. Since the main objective is to determine the combinations which have the best classification accuracy, the first twenty

combinations are shown for each technique.

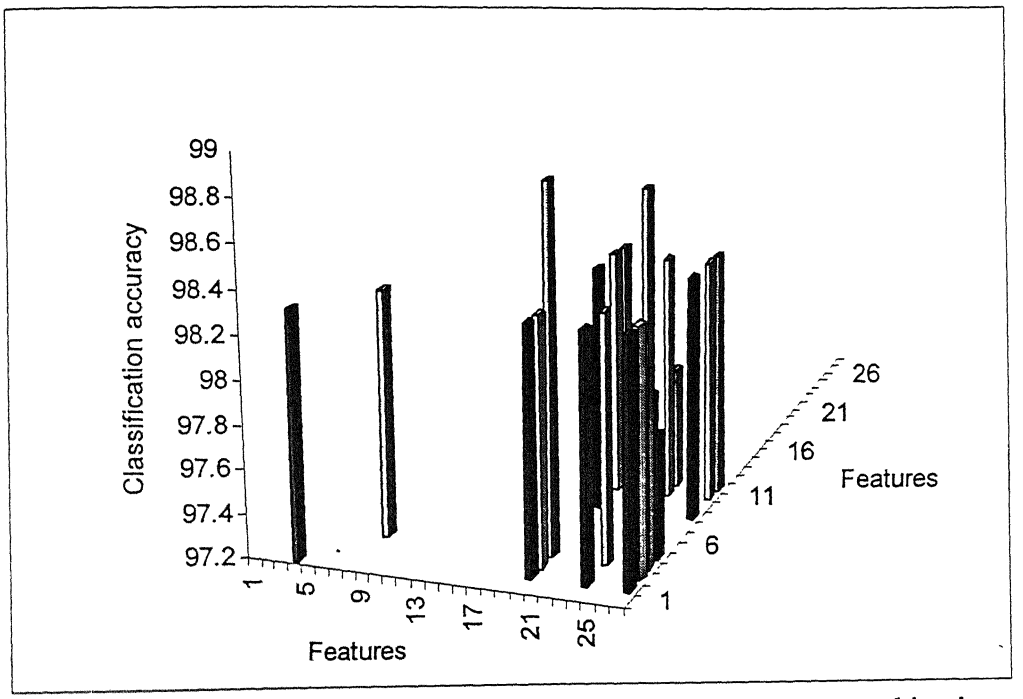


Fig. 4.8: Classification accuracy of SGLDM with two features combination

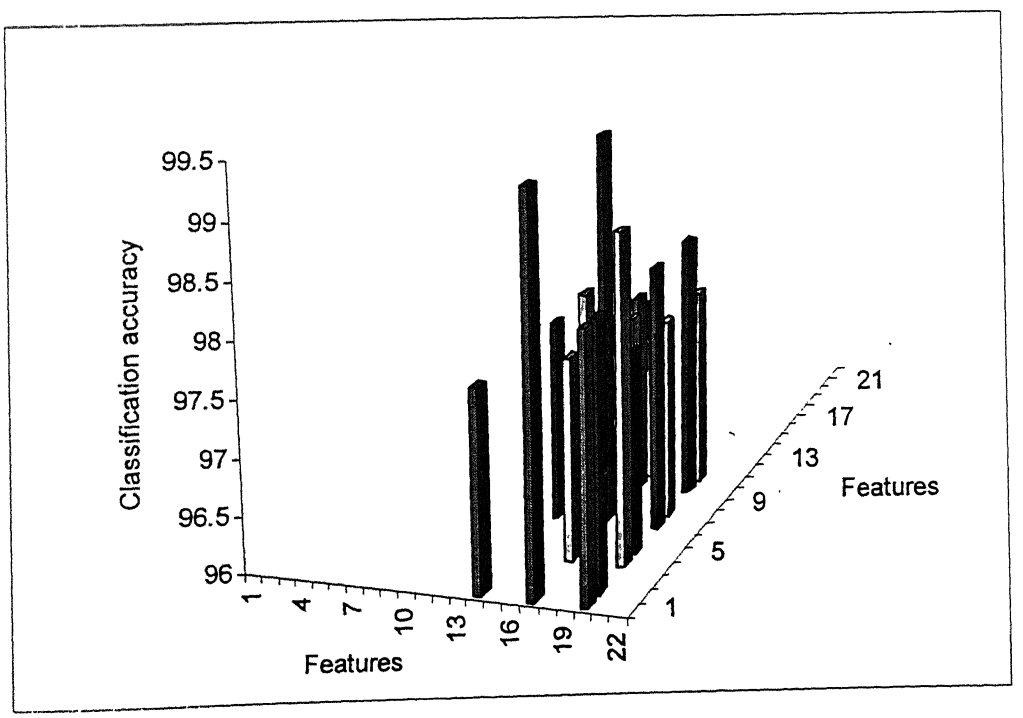


Fig. 4.9: Classification accuracy of GLDHM with two features combination

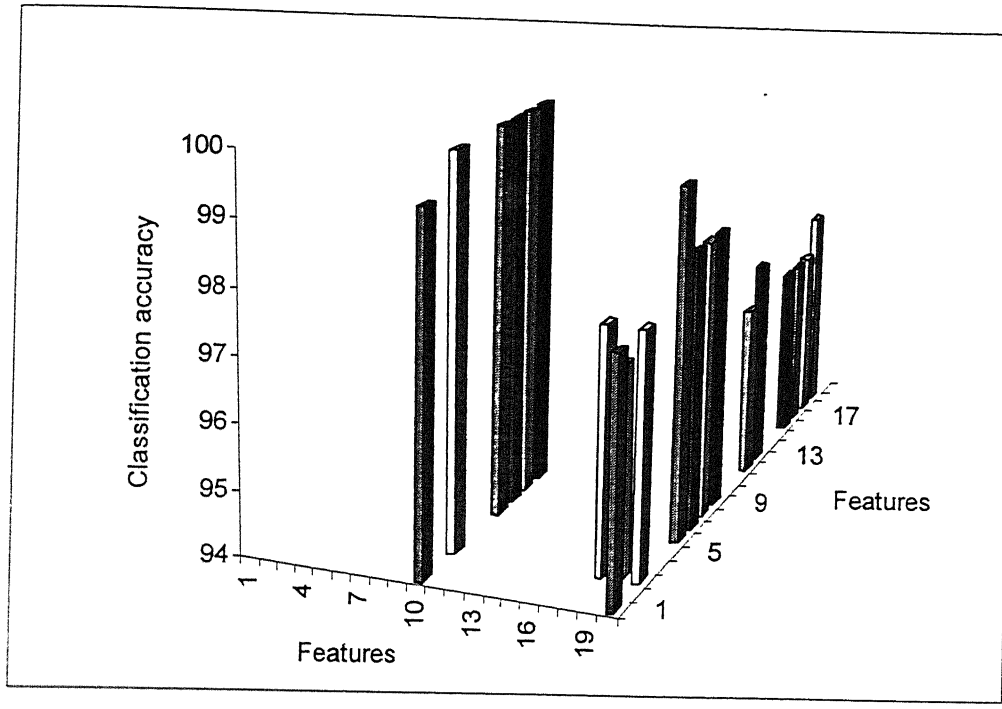


Fig. 4.10: Classification accuracy of GLRLM with two features combinations

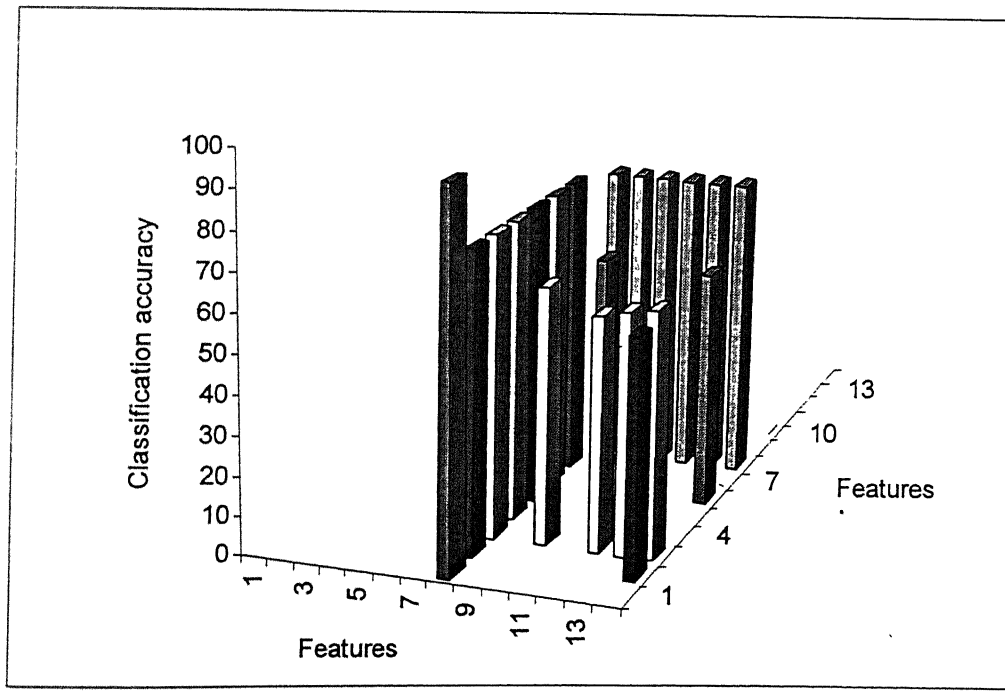


Fig. 4.11: Classification accuracy of PSM with two features combinations

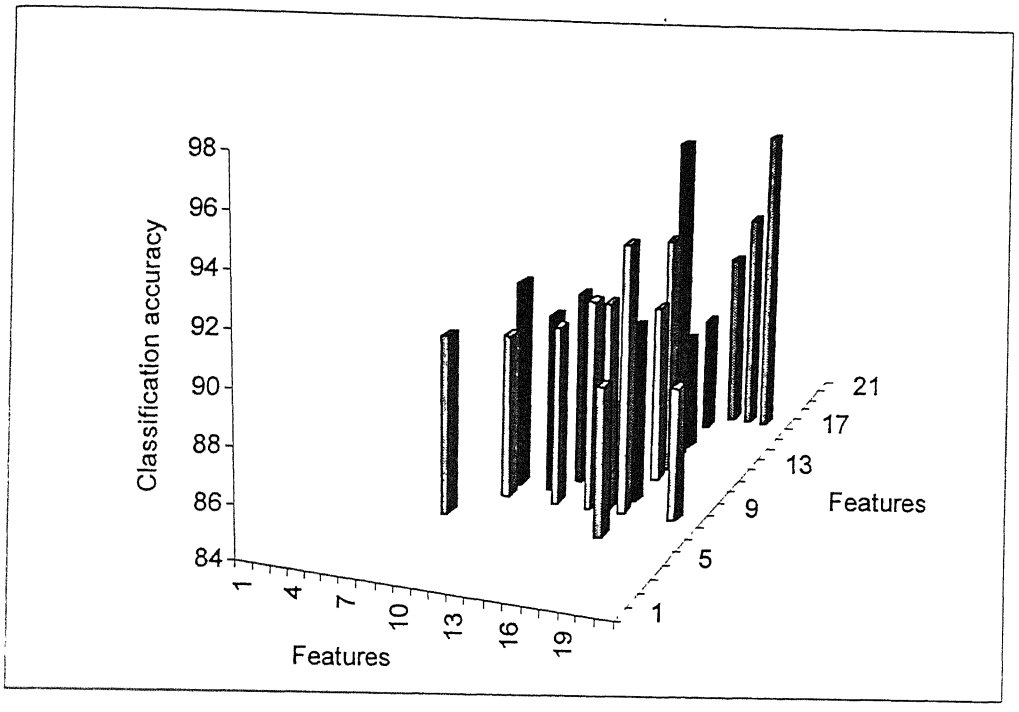


Fig. 4.12: Classification accuracy of q-φ-n with two features combinations

It is interesting to note that by considering combination of two features at a time, the classification accuracy increased close to 99% in all the methods except for PSM where it was 95% and for ϕ-q-n where it was 96%. In the case of SGLDM, the combinations of the features (4,20) and (4,27) provide a classification accuracy of 98.89%, whereas in case of GLDHM the combinations of the features (1,17) and (7,17) provide a classification accuracy of 99.44%. For GLRLM, the combinations (3,10), (6,10), (7,10), (8,10) and (9,10) provide classification accuracy of 100%. For PSM, the best combination was features (1,8) with classification accuracy of about 95%. In case of q-ϕ-n, the best combinations were features (13,17), (17,20) with classification accuracy of 96% and 95.5%, respectively. Hence, only two features are sufficient to achieve a reasonable classification accuracy. The features of best combinations can be related to the horizontal direction, horizontal-vertical

directions or vertical direction. For q-ϕ-n, the combination (17,20) is related to maximum pulse magnitude distribution. It is clear, from the above results, that there is no need to use combinations of more than two features. If more number of features are used, the computational time will increase without any substantial gain in the classification accuracy.

4.5 Conclusion

In this chapter four texture analysis algorithms were investigated for partial discharge sources classification. These techniques are spatial gray level dependence method, gray level difference histogram, gray level run length method and power spectrum method. The conventional method (q-φ-n distribution) also has been used to classify the same partial discharge sources. Minimum distance classifier has been used as supervised classifier to label unknown partial discharge source. From the work carried out in this chapter, the following conclusions can be drawn:

1. Texture analysis algorithms have high discriminative power and can be used for partial discharge sources classification. With only one feature, the classification accuracy is relatively lower (maximum 90%) which improves by considering more number of features at a time.
2. Combination of only two features are adequate for achieving high classification accuracy in all the techniques. The classification accuracy using combinations of two features are close to 99% in three of the investigated techniques.
3. These combinations consisted of features related to the horizontal direction, vertical direction and also mixed from both the directions.

4. In comparison with the conventional method (q-φ-n distribution), the texture analysis algorithms are more effective in extracting discriminating information from the partial discharge patterns especially when used to classify partial discharge patterns which are not used for the training of the minimum distance classifier.
5. The relative classification accuracy of the features depends upon the number of cycles used to construct the partial discharge patterns.
6. Power spectrum method is relatively complicated as it involves complex numbers for computation rather than real numbers. The time of calculation with this method is quite long and requires more computer memory space for storing the original and the transformed images.
7. Gray level dependence method and gray level difference histogram method do not differ much in their discriminating power. However, the one dimension gray level difference histogram method (GLDHM) performed well and yielded comparable classification accuracy with the other computationally more expensive texture calculating algorithms.

Chapter 5

Feature Selection Using Transformed Divergence Analysis

5.1 Introduction

Classification time increases with the number of features used to describe any class or partial discharge source. Removal of least effective features in classification is referred to as feature selection or feature reduction. Separability analysis is a basic approach used to identify subset of features within large number of features to facilitate the more efficient display and classification. Several measures of separability are available to predict the best features for classification[Tou and Gonzalez, 1974]. These measures are based on statistical distance between different classes. Distance between mean values of feature sets may be insufficient since overlap region is also influenced by the standard deviation of the distributions. Hence, a combination of both the distance between the mean values and a measure of standard deviation is required. Bhattacharyya distance, divergence, transformed divergence, and Jeffreys-Matusita distance can be used for this purpose. However, transformed divergence, and Jeffreys-Matusita distance have been found to have similar performance and better than the Bhattacharyya distance and the divergence measures [Mausel *et al*, 1990]. Since Jeffreys-Matusita distance is computationally more complex [Richards, 1993], the transformed divergence analysis has been used in this chapter. The main objective of this chapter is to apply the transformed divergence analysis to determine

the best feature or the best combination of more than one feature at a time which ensure the distinct separation between different partial discharge sources.

5.2 Divergence Concept

Divergence is a measure of distance or dissimilarity between two classes. It can be used to determine feature ranking and to evaluate the effectiveness of class discrimination [Tou, *et al* 1974]. Let the probability of occurrence of pattern x, given that it belongs to class w_i, be

$$P_i(x)=p(x/w_i) \tag{5.1}$$

And the probability of occurrence of pattern x, given that it belongs to class w_j, be

$$P_j(x)=p(x/w_j) \tag{5.2}$$

The discriminating information for class w_i versus class w_j may be measured by the logarithm of the likelihood ratio.

$$u_{ij} = \ln\frac{p_i(x)}{p_j(x)} \tag{5.3}$$

The average discriminating information for class w_i is then given by

$$I(i,j) = \int_x p_i(x)\times \ln\frac{p_i(x)}{p_j(x)}dx \tag{5.4}$$

The discriminating information for class w_j versus class w_i may be measured by the logarithm of the likelihood ratio

$$u_{ji} = \ln\frac{p_j(x)}{p_i(x)} \tag{5.5}$$

The average discriminating information for class w_j is then given by

$$I(j,i) = \int_x p_j(x)\times \ln\frac{p_j(x)}{p_i(x)}dx \tag{5.6}$$

The total average information for discriminating class w_i from class w_j is often referred to

as divergence, which is given by

$$J_{ij} = I(i,j) + I(j,i) \tag{5.7}$$

$$J_{ij} = \int [p_i(x) - p_j(x)] \times \ln \frac{p_i(x)}{p_j(x)} dx \tag{5.8}$$

Suppose, one has two pattern classes characterized by two n-variate normal populations

$$N(m_i, C_i) \quad and \quad N(m_j, C_j)$$

Where, m_i and m_j are the mean vectors and C_i , C_j are n × n covariance matrices. The population densities are given as,

$$p_i(x) = \frac{1}{(2\pi)^{n/2}|C_i|^{1/2}} \exp[-\frac{1}{2}(x-m_i)^t C_i^{-1}(x-m_i)] \tag{5.9}$$

and

$$p_j(x) = \frac{1}{(2\pi)^{n/2}|C_j|^{1/2}} \exp[-\frac{1}{2}(x-m_j)^t C_j^{-1}(x-m_j)] \tag{5.10}$$

From this, the logarithm of likelihood ratio can be obtained as

$$u_{ij} = \frac{1}{2}\ln\frac{|C_j|}{|C_i|} - \frac{1}{2}tr[C_i^{-1}(x-m_i)(x-m_i)^t] + \frac{1}{2}tr[C_j^{-1}(x-m_i)(x-m_i)^t] \tag{5.11}$$

The average information for discrimination between these two classes will be

$$I(i,j) = \int p_i(x_1, x_2, ..., x_n) \ln \frac{p_i(x_1, x_2, ..., x_n)}{p_j(x_1, x_2, ..., x_n)} dx_1 dx_2 ..... dx_n \tag{5.12}$$

$$I(i,j) = \frac{1}{2}\ln\frac{|C_j|}{|C_i|} - \frac{1}{2}tr[C_i(C_j^{-1} - C_i^{-1}] + \frac{1}{2}tr[C_j^{-1}(m_i - m_j)(m_i - m_j)^t] \tag{5.13}$$

Hence, the divergence for these two classes can be expressed as

$$J_{ij} = \int [p_i(x_1, x_2, ..., x_n) - p_j(x_1, x_2, ..., x_n)] \ln \frac{p_i(x_1, x_2, ..., x_n)}{p_j(x_1, x_2, ..., x_n)} dx_1 dx_2 ..... dx_n \tag{5.14}$$

or

$$J_{ij} = \frac{1}{2} tr[(C_i - C_j)(C_j^{-1} - C_i^{-1})] + \frac{1}{2} tr[(C_i^{-1} + C_j^{-1})(m_i - m_j)(m_i - m_j)^t] \quad (5.15)$$

where, *tr* is the trace of the matrix, and the subscript t denotes the transpose of the matrix. The divergence possesses the following useful properties [Tou and Gonzalez, 1974]:

1. $J_{ij} > 0$ for $i \neq j$

2. $J_{ij} = 0$ for $i = j$

3. $J_{ij} = J_{ji}$

4. J_{ij} is additive for independent features

$$J_{ij}(x_1, x_2, \ldots\ldots\ldots x_m) = \sum_{k=1}^{m} J_{ij}(x_k) \quad (5.16)$$

5. Adding a new feature never decreases the divergence

$$J_{ij}(x_1, x_2, \ldots\ldots\ldots, x_m) \leq J_{i,}(x_1, x_2, \ldots\ldots\ldots, x_m, x_{m-1}) \quad (5.17)$$

The additive property of divergence implies that, if the features are independent, the divergence based on *m* features is equal to the sum of the *m* divergences based on each feature separately. This property may be utilized to determine the relative importance of each of the features to be selected. The features which will lead to large divergence between the classes are more important, since they carry more discriminatory information. Thus one may rank the importance of each feature according to its associated divergence. However the use of divergence may lead to the following problem in some cases.

If divergence is plotted versus distance between classes it increases quadratically. This behavior, unfortunately, is quite misleading if divergence is to be used as an indicator of how successfully patterns could be discriminated or classified. It implies that at large separations, further increase in the separation will lead to better classification accuracy,

whereas in practice this is not the case in probability of correct classification. Moreover, easily separable classes will weight average divergence upward in a misleading fashion to the extent that sub-optimal reduced features subset might be indicated as the best. This drawback of divergence method can be overcome by using the transformed divergence [Richards, 1993] described in the next section.

5.3 Transformed Divergence

A useful modification of divergence to over come its drawback is to use the transformed divergence. The transformed divergence is given by

$$J^T(i,j) = 2(1 - e^{-J(i,j)/8}) \qquad (5.18)$$

The presence of exponential factor gives an exponentially decreasing weight to increasing separation between the partial discharge classes similar to that expected for the probability of correct classification. In other words, it will have a saturating behavior with increasing separation of the classes. It is asymptotic to 2.0. Thus a transformed divergence of 2.0 between partial discharge sources would imply that the classes of these partial discharge sources can be identified with 100% classification accuracy. This saturating behavior is highly desirable since it does not suffer the difficulty experienced with divergence method.

5.4 Use of Transformed Divergence for Feature Selection

Consider a case of M partial discharge sources (classes), each having total N features, and there is a need to select the best n features subset. The transformed divergence between each pair of classes have been determined for all combinations of n features out of the N features. An average indication of separability is then given by computing the average transformed divergence.

$$J^{T}{}_{ave} = \sum_{i=1}^{M}\sum_{j=1}^{M} p(w_i)p(w_j)J_{ij} \qquad (5.19)$$

where $p(w_i)$ and $p(w_j)$ are the class prior probabilities.

5.5 Results and Discussion

In this study, six different partial discharge sources have been used which were created in the laboratory as described in chapter-3. The four texture analysis algorithms, gray level difference histogram method, spatial gray level dependence method, gray level run length and the power spectrum method as well as the q-φ-n distribution have been used to generate the features used to calculate the average transformed divergence. Using each feature individually of these algorithms, the transformed divergence between every pair of classes was calculated. The average transformed divergence between the six classes for each feature of the four texture analysis algorithms and the q-φ-n distributions is given in the following sub-sections.

5.5.1 Grey level difference histogram method

Gray level difference histogram has five features (f1, f2,....,f5) plus the feature of polarity factor (f6) as described in chapter 2. Using these features for horizontal and vertical direction of the positive and negative half cycles, the average transformed divergence between the six partial discharge sources is shown in Fig. 5.1 for different number of cycles. Fig. 5.2 shows the effect of changing the number of cycles on the discriminating power of the features when used in the horizontal and vertical directions of the positive and negative half cycles, respectively. The number of cycles were 2, 4, 8, 16, 32 as given in Fig. 5.2 and denoted as A, B, C, D, E, respectively. In this figure, the symbols A, B, C, D,

E were repeated four times, the first one for the features applied on the horizontal direction of the positive half cycle, the second one for the features when applied for the vertical direction of the positive half cycle while the third and fourth for the features when applied for the horizontal and vertical directions of the negative have cycles, respectively. From figures 5.1 and 5.2 it can be noticed that:

(i) Feature number 17 has the maximum separation between the partial discharge sources with average transformed divergence values of 1.99, 1.99, 1.88,1.83 and 1.75 for 2, 4, 8, 16 and 32 cycles, respectively (Fig. 5. 1).

(ii) While the discriminating power of feature number 17 is decreasing with increasing number of cycles, it still has the maximum discriminating power compared to the other features except in the case of 32 cycles (Fig. 5. 1).

(iii) In the positive half cycles, the features describing the horizontal direction have almost the same average transformed divergence as the features describing the vertical direction. For a given direction and number of cycles, the features f1, f2, f4 and f5 have the same average transformed divergence (Fig. 5.2).

(iv) In the negative half cycle, for a given direction and a number of cycle, the features f2, f4, f5 have the same average transformed divergence (Fig. 5. 2).

(v) In the same half cycle, increasing the number of cycles has the same effect on the features used with the horizontal and the vertical directions (Fig. 5.2).

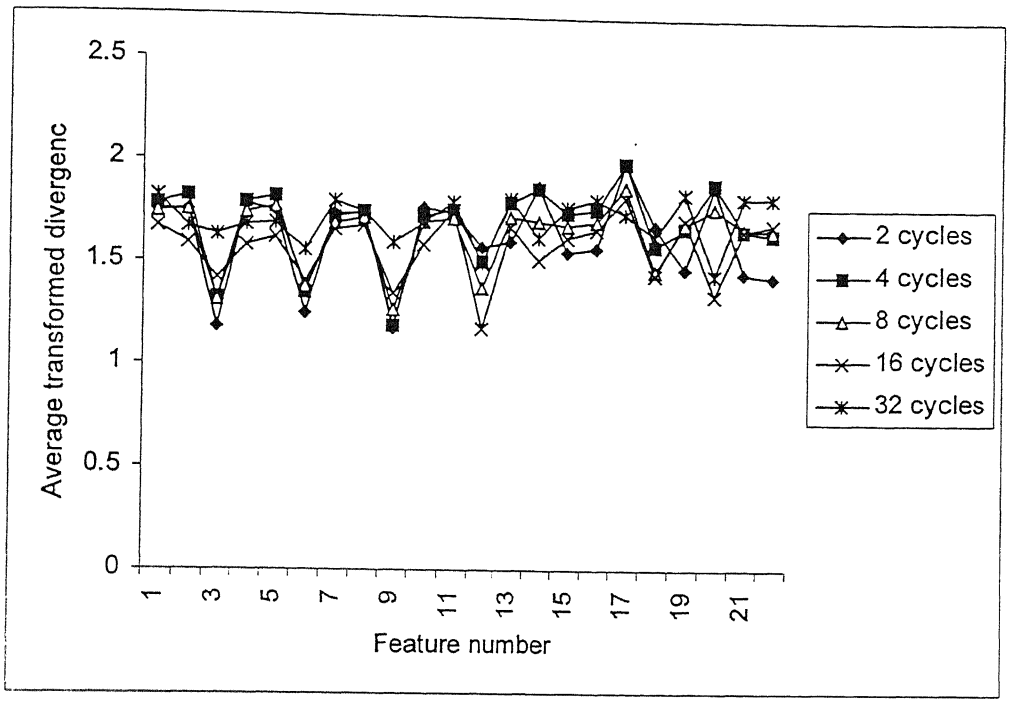


Fig. 5.1: Average transformed divergence between the six partial discharge sources using the GLDHM features

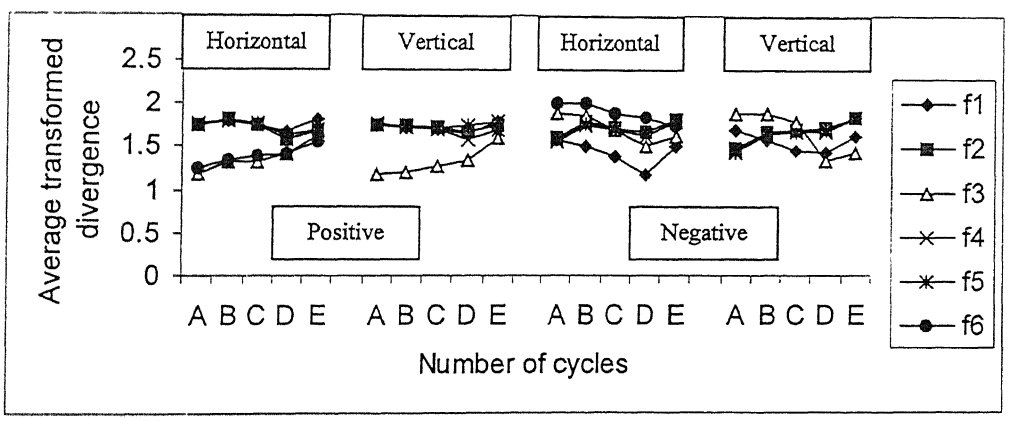


Fig. 5.2 : Effect of the number of cycles on the average transformed divergence between the six partial discharge sources using the GLDHM features
(A-2 cycles, B-4 cycles, C-8 cycles, D-16 cycles, E-32 cycles)

5.5.2 Spatial gray level dependence method

Spatial gray level dependence method has seven features (f1, f2,....,f7) as described in chapter 2. The transformed divergence of these features, when applied for the positive and negative half cycles in the horizontal and vertical directions, are shown in Fig. 5.3. The

effect of increasing the number of cycles on the discriminating power of these features when used in the horizontal and vertical directions of the positive and negative half cycles, respectively, is shown in Fig. 5.4. From these figures, it can be observed that:

(i) Correlation between the partial discharge pulses using the vertical direction of positive half cycle (feature number 14) and negative half cycles (feature number 28) has the worst discriminating power (Fig. 5.3).

(ii) Except the correlation, in each half cycle, the effect of increasing the number of cycles on the relative discriminating power is the same in the horizontal and vertical directions (Fig. 5.4).

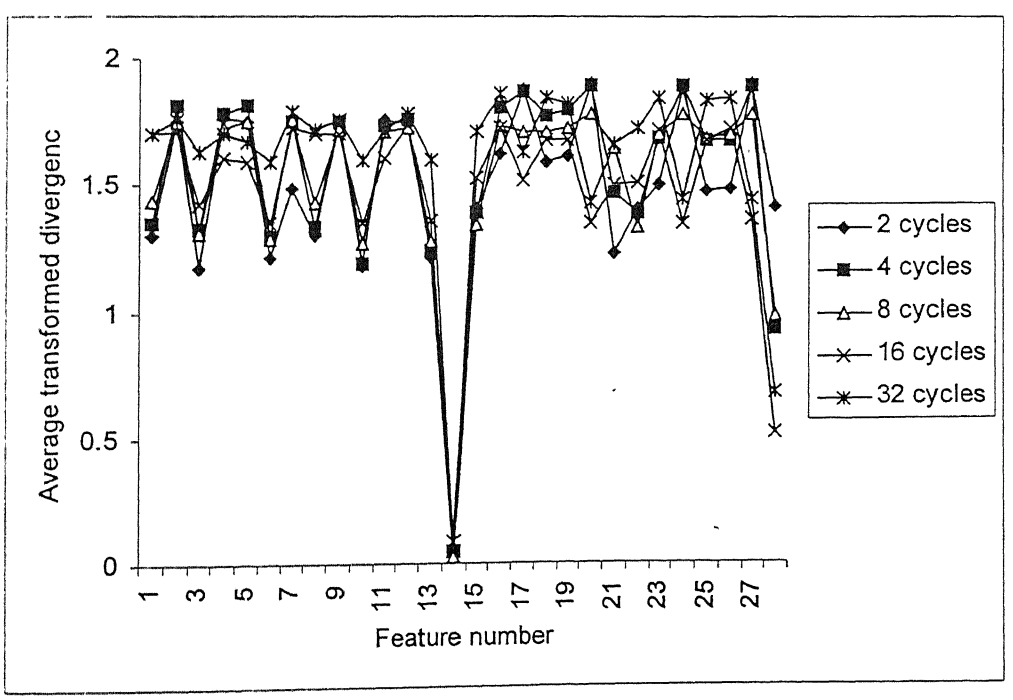


Fig. 5.3 : Average transformed divergence between the six partial discharge sources using the SGLDM features

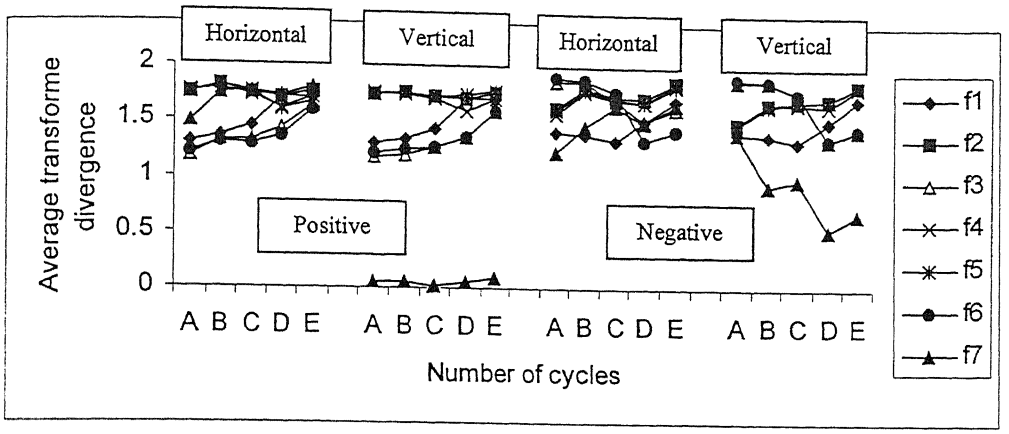


Fig. 5.4 : Effect of the number of cycles on the average transformed divergence between the six partial discharge sources using the SGLDM features (A-2 cycles, B-4 cycles, C- 8 cycles, D-16 cycles, E-32 cycles)

(iii) In both the directions of the positive half cycle, the discriminating power of the features (f1, f3, f6) increases with increasing the number of cycles (Fig. 5.4).

(iv) In each half cycle, the contrast (f3) and the variance (f6) has the same discriminating power in both the directions (Fig. 5.4).

(v) For a given direction and number of cycles, the angular second moment (f2), the entropy (f4) and the IDM (f5) have the same discriminating power (Fig. 5.4)

5.5.3 Gray level run length method

The average transformed divergence between the six partial discharge sources using the five features of the gray level run length method (f1, f2,...,f5), as described in chapter 2, is shown in Fig. 5. Fig 6 shows the effect of increasing the number of cycles on the average transformed divergence of the gray level run length method features when used in the horizontal and vertical directions of the positive and negative half cycles, respectively.

From these figures, the following observations can be made:

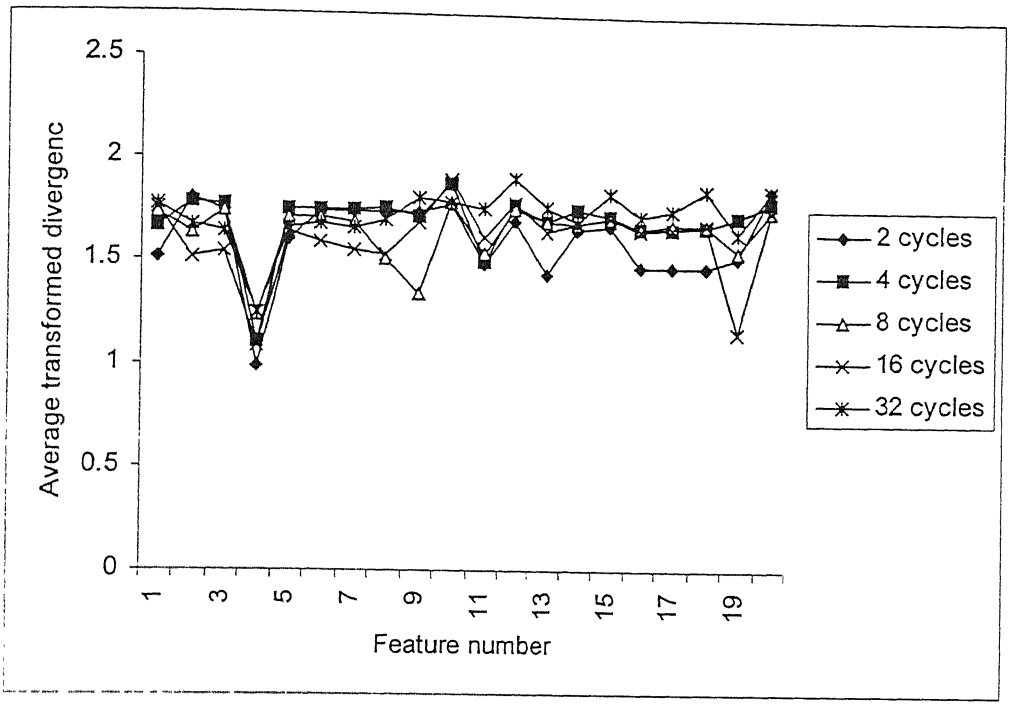


Fig. 5.5 : Average transformed divergence between the six partial discharge sources using the GLRLM features

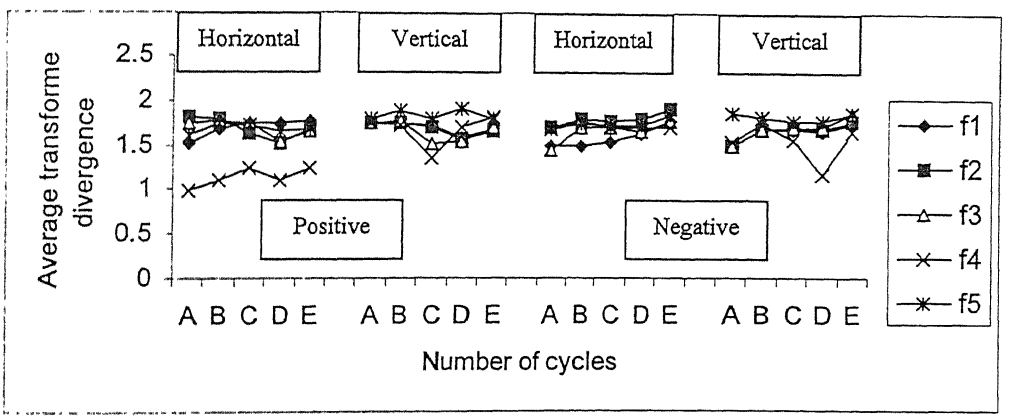


Fig. 5.6 : Effect of the number of cycles on the average transformed divergence between the six partial discharge sources using the GLRLM features (A-2 cycles, B-4 cycles, C-8 cycles, D-16 cycles, E-32 cycles)

(i) Run length non-uniformity (f4) has the worst discriminating power when used in the horizontal direction of the positive half cycle (Fig. 5.5).

(ii) Gray level non-uniformity (f3) has the same discriminating power when

used in the horizontal and vertical directions of the negative half cycle (Fig. 5.6).

(iii) The run percentage (f5) has the best discriminating power when used in the vertical direction of any half cycle irrespective of the number of cycles used (Fig. 5.6).

5.5.4 Power spectrum method

The number of features, in this method, depends upon the number of power frequency cycles used to construct the partial discharge patterns as shown in Fig. 5.7 and as described in chapter 2. As shown in Fig. 5.8, for a given polarity (positive or negative half cycle) and number of cycles, the features derived from the horizontal direction have the same discriminating power. From Fig. 5.8, the following observations can be made:

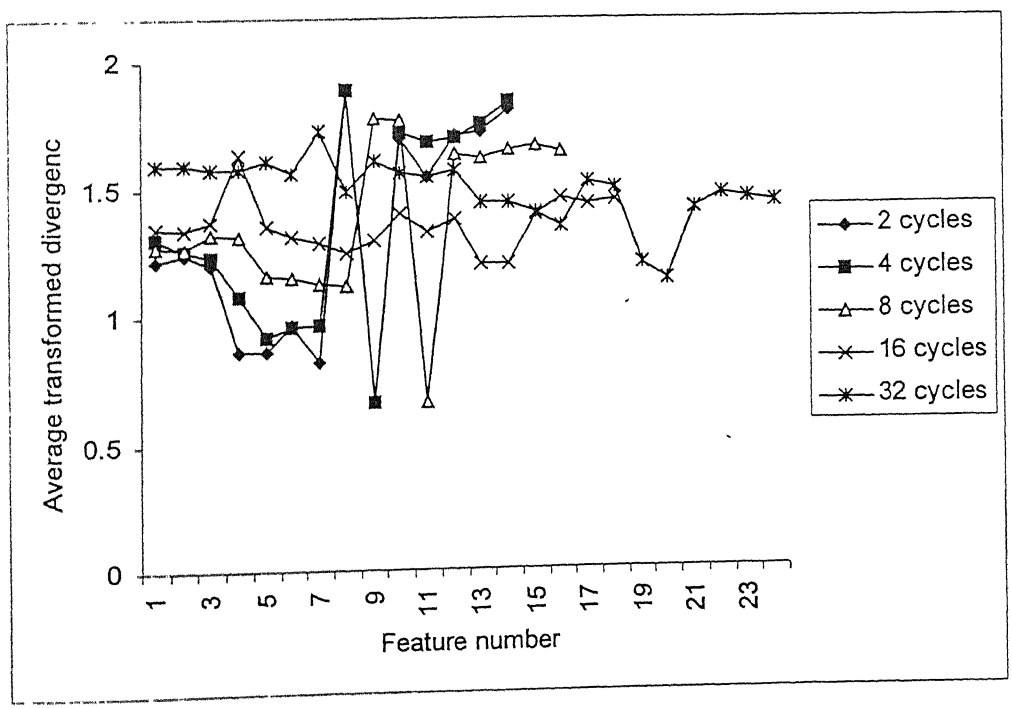


Fig. 5.7 : Average transformed divergence between the six partial discharge sources using the PSM features

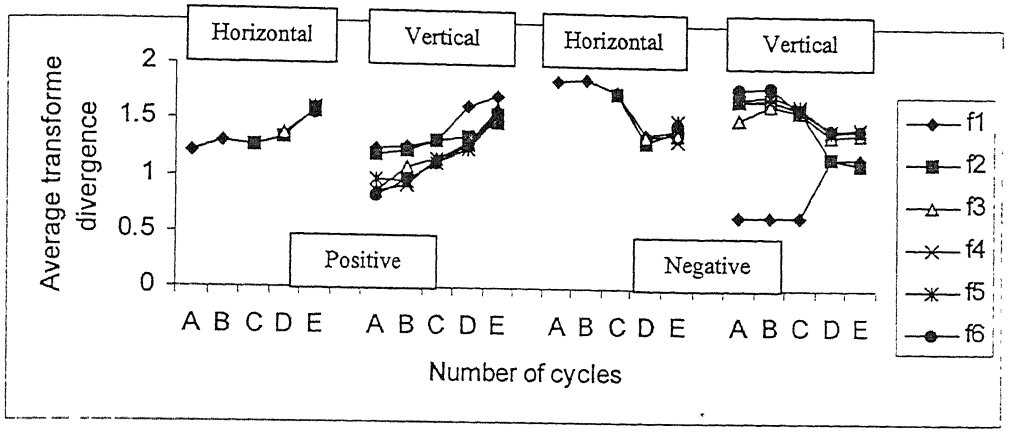


Fig. 5.8 : Effect of the number of cycles on the average transformed divergence between the six partial discharge sources using the PSM features (A-2 cycles, B-4 cycles, C- 8 cycles, D-16 cycles, E-32 cycles)

(i) Increasing the number of power frequency cycles improves the discriminating power of the features derived from the horizontal or vertical directions in the positive half cycle. However, an opposite effect has been noticed in the negative half cycle (Fig. 5.8).

(ii) With lower number of cycles, the features derived in both the horizontal and vertical directions in the negative half cycle are much better as compared to the corresponding features in the positive half cycle (Fig. 5.8).

(iii) Feature number 6 has the best discriminating power when derived from the vertical direction of the negative half cycle. However, its effect is very poor when derived from the vertical direction of the positive half cycle (Fig. 5.8).

(iv) Feature number 1 has a very poor discriminating power when derived from the vertical direction of the negative half cycle (Fig. 5.8).

5.5.5 q-φ-n distributions method

From the three dimensional distribution q-φ-n, three two dimensional distributions have been derived. These distributions are number of pulses vs. phase angle (n-φ), average partial discharge pulse magnitude vs. phase angle (ave.-φ) and the maximum partial discharge pulse magnitude vs. phase angle (max.- φ). Each distribution has seven features (f1, f2,.....,f7) as described in chapter 1. Fig. 5. 9 shows the discriminating power of the different features for different number of cycles and Fig. 5.10 shows the effect of increasing the number of power frequency cycles on the discriminating power of these features. The letters A, B, C, D and E, which correspond to the number of cycles used (2, 4, 8, 16, 32 cycles, respectively), were repeated three times corresponding to the number of distributions mentioned earlier. From Figs. 5.9 and 5.10, the following observations can be made:

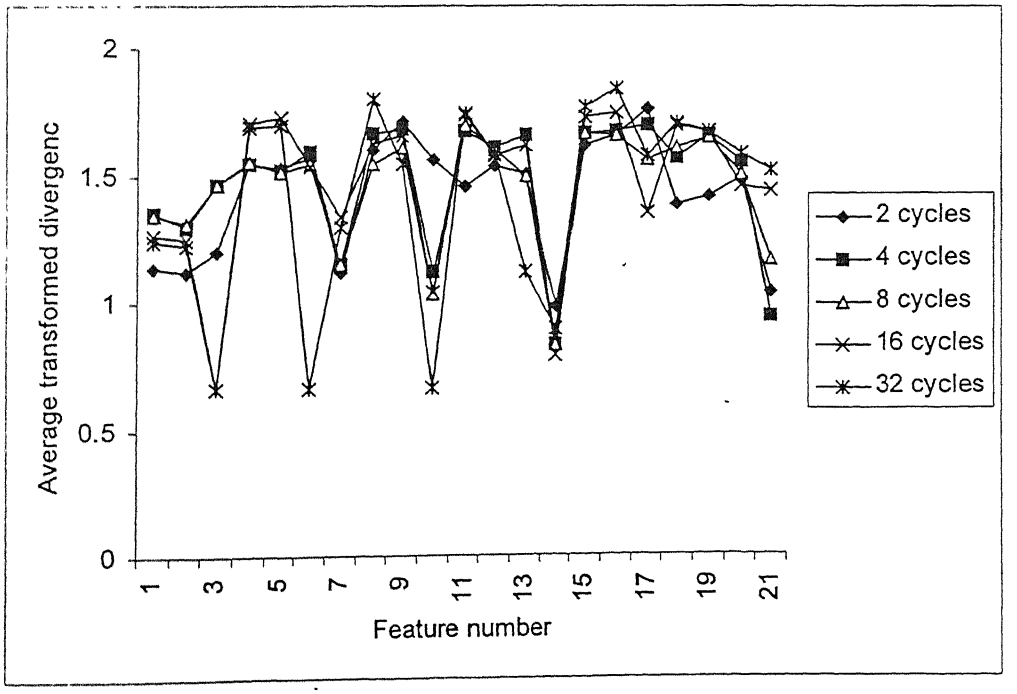


Fig. 5.9 : Average transformed divergence between the six partial discharge sources using the q-φ-n distributions features

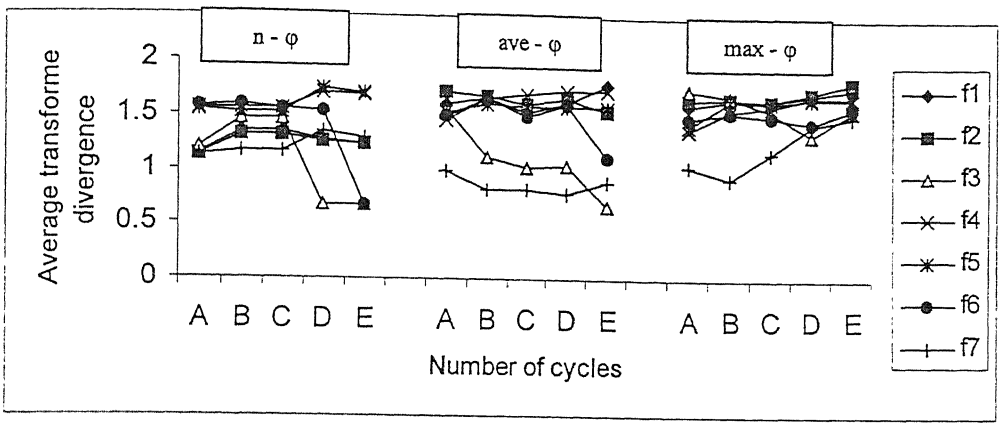


Fig. 5.10 : Effect of the number of cycles on the average transformed divergence between the six partial discharge sources using the q-φ-n distributions features (A-2 cycles, B-4 cycles, C- 8 cycles, D-16 cycles, E-32 cycles)

(i) The correlation between the partial discharge pulses in the positive and negative half cycles (f7 in Fig. 5.10) has poor discriminating power in the second distributions (feature 14 in Fig. 5.9) compared with the correlation in the first and third distributions (feature 7, 21 in Fig. 5.9).

(ii) The features derived from the third distribution is better than the features derived from the other distributions (Fig. 5.9).

(iii) In the positive half cycle of the first distribution, skewness (feature 1 in Fig.5.9) is better than kourtosis (feature 2 in Fig. 5.9). They have almost the same discriminating power in the other distributions (features 8, 9 and 15, 16 in Fig. 5.9).

5.5.6 Effect of increasing the number of features used for measuring the separation

All the above results were obtained using one feature at a time. From these results, the maximum average transformed divergence between the classes at different number of power frequency cycles were found as given in table 5.1.

Table 5.1: Maximum average transformed divergence obtained for the different algorithms at different number of power frequency cycles.

	Number of cycles				
	2	4	8	16	32
Max TD GLDHM	1.999	1.99	1.88	1.83	1.84
Max TD SGLDM	1.9	1.89	1.78	1.73	1.86
Max TD GLRLM	1.86	1.88	1.79	1.90	1.91
Max TD PSM	1.87	1.88	1.77	1.63	1.72
Max TD q-φ-n	1.75	1.69	1.69	1.74	1.83

The above table reveals the following:

(i) The features of GLDHM have a good discriminating power for the different number of cycles in comparison with the other techniques.

(ii) The features of q-φ-n distributions have relatively poor discriminating power compared to the features of GLDHM .

(iii) More than one feature has to be used at a time to improve the separability between the different classes.

Hence, combination of two features were used at a time to find the best combination which has the maximum discriminating power. The average transformed divergence between the six partial discharge sources were investigated using all possible combinations. The main objectives have been to:

- check if two features are sufficient for complete separation between the classes.
- investigate the relationship between the direction of features of each combination in order to determine the direction (horizontal or vertical) providing combinations.

To meet the above objectives, the number of combinations having average transformed divergence greater than certain values were determined. These values were 2, 1.95, 1.9, 1.85, 1.8, 1.75, 1.7, 1.65, 1.6, 1.55 and 1.5. The combinations greater than any one of these values were divided into three categories:

- Horizontal (Hor.), if both the features were related to the horizontal direction.
- Vertical (Ver.), if both the features were related to the vertical direction only.
- Mixed (Mix.), if one feature was related to the horizontal direction while the second one was related to the vertical direction.

For each texture algorithm, the number of combinations in the horizontal, vertical or the mixed directions is given as percentage of the total number of possible combinations for that algorithm. Figs. 5.11a to 5.11e show the distribution of the combinations for the texture analysis algorithms in addition to the q-φ-n distributions. In each figure, the values 2, 1.95,.....,1.5 have repeated 5 times corresponding to the number of power frequency cycles used i.e. 2, 4, 8, 16 and 32 cycles.

Fig. 5.11a shows the changes in the percentage of combinations, for the different number of cycles, with changing the value of the average transformed divergence in the case of GLDHM. From this figure, it can be noticed that the combinations from horizontal direction only is greater than the combinations from the vertical direction. The difference between them is almost constant.

Fig. 5.11b shows the results with the SGLDM. The combinations related to the horizontal direction are greater than the combinations related to the vertical direction at the higher values of the average transformed divergence values. With decreasing the average transformed divergence, the percentage of the combinations related to the horizontal and vertical directions become the same.

Fig. 5.11c shows the results with GLRLM. In this case, the percentage of the combinations in the horizontal and the vertical directions are the same with the different values of average transformed divergence as well as with changing the number of cycles.

In the case of PSM, as shown in Fig. 5.11d, it was noticed that there is no combination of features related to the horizontal direction at the lower values of the number of cycles. However, with increasing the number of cycles the combinations from the horizontal direction become comparable to those from the vertical direction.

In case of q-φ-n distributions, the combinations were divided into six categories consisting of combinations from (n-φ) only, combinations from (n-φ) and (ave.-φ), combinations from (n-φ) and (max.- φ), combinations from (ave.-φ) only, combinations from (ave.-φ) and (max.- φ) and combinations from (max.- φ) only. From Fig. 5. 11e, it can be observed that the feature combinations from first distribution (n-φ) has poor discriminating power compared to the other distributions. Also, it was noticed that the percentage of combinations from different distributions is better than the combinations of any single distribution.

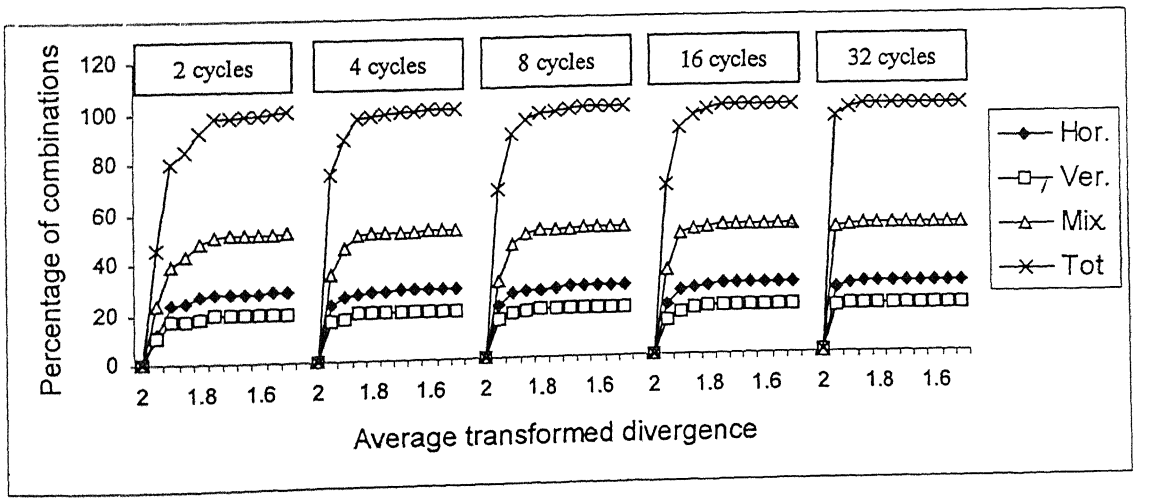


Fig. 5.11a: GLDHM

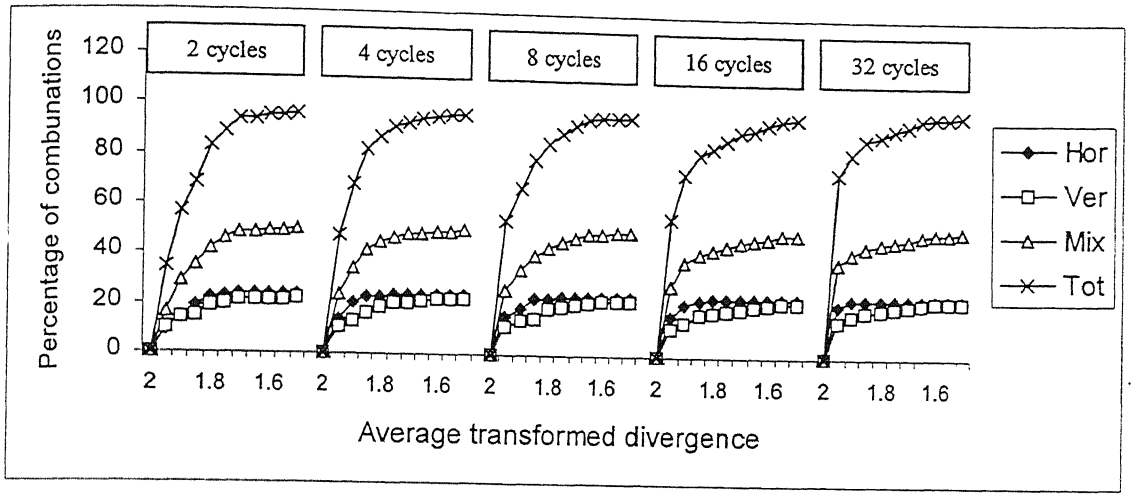


Fig. 5.11b: SGLDM

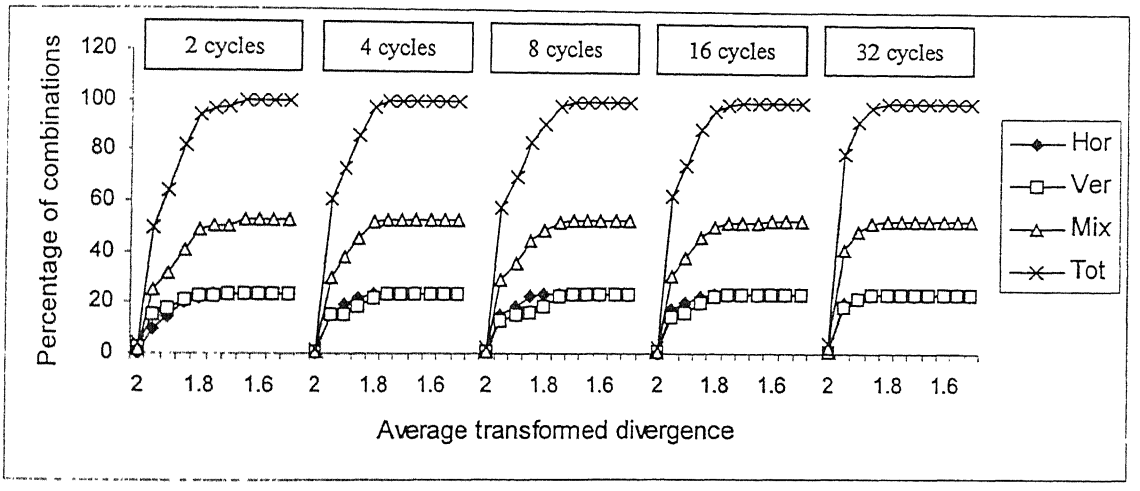


Fig. 5.11c: GLRLM

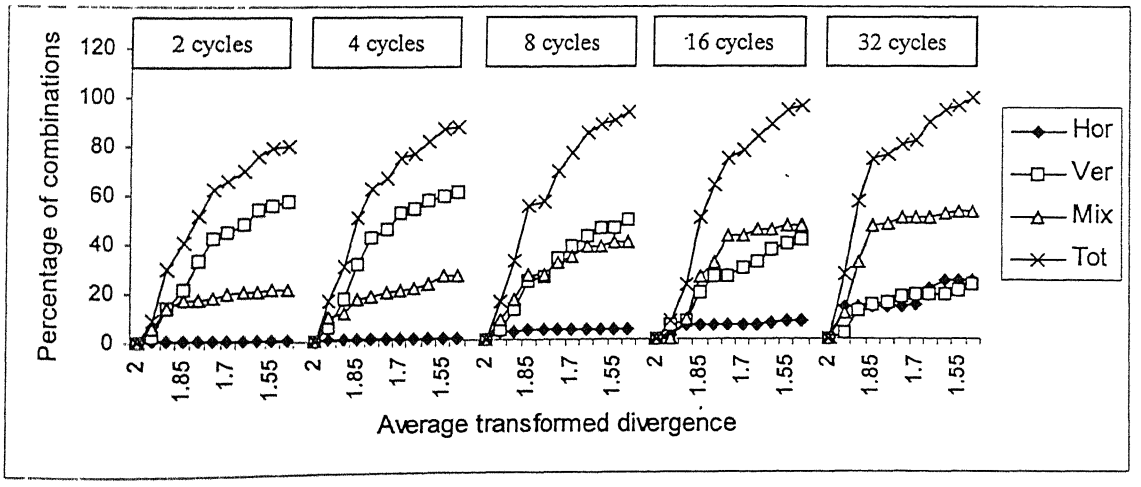


Fig. 5.11d: PSM

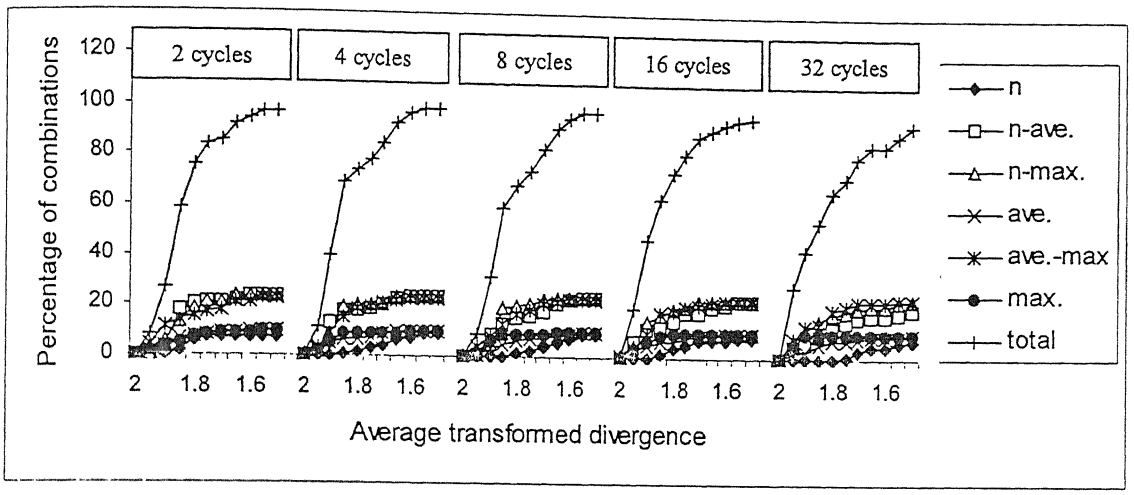


Fig. 5.11e: q-φ-n distributions

Fig. 5.11: The relation between the features belonging to the combinations and the direction used to generate these features for different number of cycles and different average transformed divergences

Thus, from Fig. 5. 11, the following general observations can be made:

(i) In all the different techniques, there are some combinations which have an average transformed divergence equal to 2.

(ii) In the case of GLDHM, SGLDM and GLRLM, the number of combinations from the horizontal direction are almost equal to the combinations from the vertical direction. Also, the number of combinations from mixed features have almost the same value for different number of cycles.

(iii) In the case of the four texture analysis algorithms, the mixed combinations are almost equal to the summation of the horizontal and vertical combinations.

(iv) With increasing the number of cycles, the total number of combinations ater than a certain average transformed divergence, increases. In other

words, the curve of the total number of combinations reach the level of 100% at higher values of the average transformed divergence.

(v) The combinations of GLDHM, SGLDM and GLRLM methods are much better than the PSM and q-φ-n distributions. The percentage of the total number of combinations were 100%, 95% and 100% at average transformed divergences of 1.85, 1.5, 1.8, respectively in the case of 32 cycles.

5.6 Comparison of Results With Divergence Method

For comparing the results, the average divergence between the six partial discharge sources was calculated using the GLDHM only as shown in Fig. 5.12. From this figure, one can notice that the divergence has no limit. It indicates that most of the features used for the negative half cycles has more average divergence compared to the features used for the positive half cycle. Also it is giving a misleading result that the feature number 17 (which is the best feature as shown in Fig. 5.1) has no discriminating power compared to the other features. At the same time it indicate that, for example, feature number 18 has better discriminating power compared to feature number 17. Actually, feature number 17 has better discriminating power compared to feature number 18 as shown in Figs 5.13, 5.14 which give the distributions of the different partial discharge classes using features number 17 and 18. In these figures, 180 patterns were used consisting of 30 patterns for each class. As shown in Fig 5.13, all the classes are concentrated around its mean values and there is no overlapping between the classes 1, 3, 4, 5 and 6. On the other hand, by using feature number 18 there is overlapping between classes 1,2 and 2,3,4 as well as 3,6.

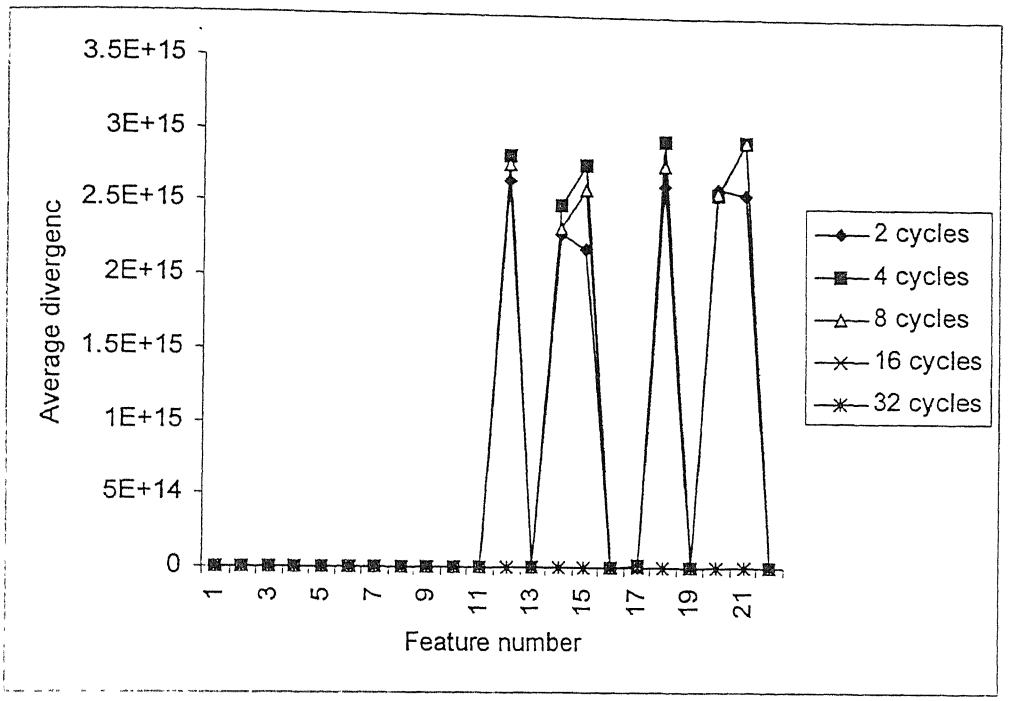


Fig. 5.12: Average divergence between the six partial discharge sources using the GLDHM features

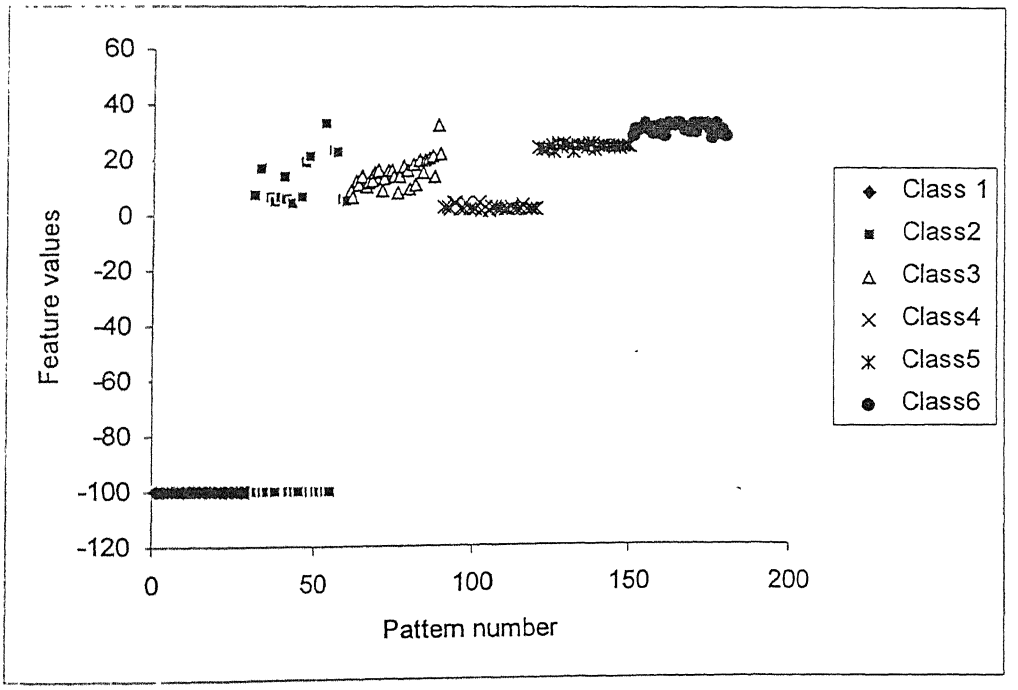


Fig. 5.13: Distribution of the patterns of the six partial discharge sources using feature number 17

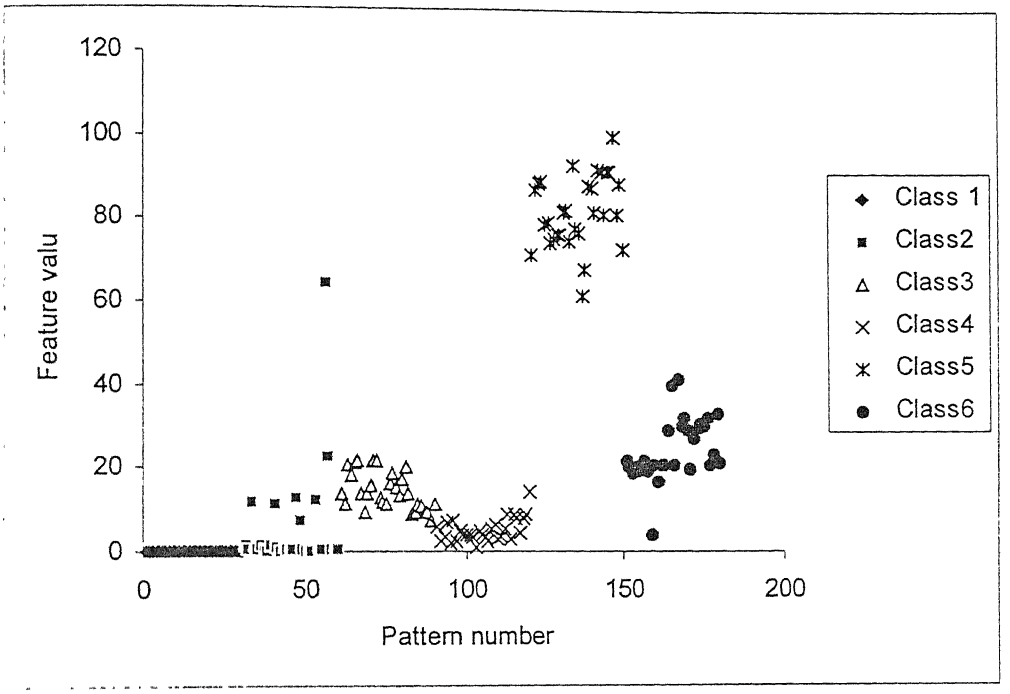


Fig. 5.14: Distribution of the patterns of the six partial discharge sources using feature number 18

Tables 5.2 and 5.3 show the divergence and the transformed divergence between the partial discharge sources using feature number 17 while tables 5.4 and 5.5 show the divergence and the transformed divergence between the partial discharge sources using feature number 18. Comparing tables 5.2 and 5.4, it can be noticed that, except the divergence between the first class and the other classes, feature number 17 has better divergence between the classes compared to feature number 18. But the average divergence between the classes using features 17 and 18 are 9.68E+12 and 277E+15, respectively implying that feature number 18 is better than feature number 17. This means that if the divergence between only two classes, using a certain feature, is very high, the average divergence between all the classes will be high and this feature will be considered as the best irrespective of the divergence between the other classes. It should be noted that the low value of the standard deviation of class 1 is the reason of the high value of divergence between that class and the

other classes.

On the other hand by comparing tables 5.3 and 5.5, it was noticed that the transformed divergence between the classes using feature number 17 is much better than the transformed divergence using feature number 18. The average transformed divergence between the classes using features 17 and 18 were 1.99 and 1.68, respectively. From the above, it is clear that the average transformed divergence is more reasonable when used for the features selection compared to the divergence.

Table 5.2: Divergence between six partial discharge sources using feature number 17 of GLDHM

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class 1	0	1.52E+13	2.75E+13	2.19E+13	3.23E+13	3.59E+13
Class 2		0	104.5221	3482.655	5628.897	1377.131
Class 3			0	133.1455	100.7227	56.13154
Class 4				0	752.3771	749.5209
Class 5					0	44.29177
Class 6						0

Table 5.3: Transformed divergence between six partial discharge sources using feature number 17 of GLDHM

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class 1	0	2	2	2	2	2
Class 2		0	1.999996	2	2	2
Class 3			0	2	1.999993	1.998206
Class 4				0	2	2
Class 5					0	1.992119
Class 6						0

Table 5.4: Divergence between six partial discharge sources using feature number 18 of GLDHM

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class 1	0	9.11E+14	1.13E+15	1.69E+14	3.40E+16	3.25E+15
Class 2		0	5.555839	9.008136	58.90788	5.043419
Class 3			0	8.136919	149.3223	4.092518
Class 4				0	421.3721	29.73506
Class 5					0	51.7099
Class 6					0	0

Table 5.5: Transformed divergence between six partial discharge sources using feature number 18 of GLDHM

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class 1	0	2	2	2	2	2
Class 2		0	1.001332	1.351355	1.998732	0.935272
Class 3			0	1.276726	2	0.800887
Class 4				0	2	1.951381
Class 5					0	1.996882
Class 6						0

5.7 Conclusion

In this chapter the average transformed divergence has been used to measure the separability between the different partial discharge sources. From the results achieved in this chapter, it was found that:

(i) The GLDHM features have a good discriminating power between different partial discharge sources in comparison with the other techniques.

(ii) The combinations which have features related to the horizontal direction are equal or greater than the combinations of features related to the vertical directions.

(iii) The optimum features which give high separation between the different classes, according to the transformed divergence analysis (these features are supposed to give maximum classification accuracy), are slightly different than those which are determined according to the minimum distance classification accuracy given in chapter 4. For example, the correlation (feature number 7) of the SGLDM, when used with the horizontal direction of the positive half cycle the minimum distance classifier gives the maximum classification accuracy. However, when used with the

transformed divergence, it was not the best. It is suspected that the non-Gaussian distribution of some classes is the cause of the disagreement between the classification accuracy results and the transformed divergence analysis results.

(iv) By using the transformed divergence, the effect of changing the number of cycles on the discriminating power of the features were slightly reduced compared with the result of the minimum distance classifier.

Chapter 6

Neural Network Based Partial Discharge Patterns Classification

6.1 Introduction

In chapter 4, minimum distance classifier has been used for partial discharge sources classification by using features based on four texture analysis techniques which had offer better classification accuracy as compared to the conventional q-φ-n method. To use the minimum distance classifier, like any other statistical classifier, various preliminary conditions e.g. data from normal populations, must be fulfilled in order to carry out the analysis [Tou and Gonzalez, 1974]. However, it was observed that some partial discharge sources generate classes that deviate from the normal distributions. This deviation creates a problem with minimum distance classifier as well as with features selection techniques like the transformed divergence analysis. On the other hand the artificial neural networks (ANNs) approach basically belongs to the nonparametric methods. Hence, it is not necessary to make any assumption about data structure. Also ANNs have the ability to form non linear decision boundaries between the different classes in the feature space. Although the field of ANNs is now a well established area of endeavor, its application to partial discharge is relatively very recent. Some of the applications of ANN in the field of partial discharge sources classification are as following.

A three layer feed forward ANN with back propagation learning algorithm has been used for the automatic discrimination of partial discharge, generated in an XLPE cable, from noise (Suzuki *et al*, 1992). Three types of input patterns; ϕ–*q-1*, ϕ–*q-n*, and ϕ–*q-n`* were used. It was noticed that the ANN could easily discriminate partial discharge from noise by using ϕ–*q-1* patterns.

ANN using back propagation method has been applied to the discrimination of partial discharge patterns before and after the tree initiation from a needle-shaped void [Hozumi *et al*, 1992]. The input patterns were based on the ϕ–*q-n* and ϕ–*q* distributions. Similar technique with the same input patterns was used to classify seven different partial discharge sources by using three dimensional distributions [Satish et al, 1994].

Another approach based on 15 statistical parameters like skewness, kurtosis, number of peaks and cross correlation had been used to define input patterns of the ANN [Gulski et al, 1993]. Back propagation network, Kohonen self organising map and learning vector quantization network were used. The back propagation network provided satisfactory results for all the studied partial discharge sources compared to the other types of ANNs.

Another method which was not based on these distributions but focused on features that describe the shape of the partial discharge pulses, namely the apparent charge, rise time, fall time, width and area of the partial discharge pulse had also been used [Mazroua *et al*, 1993]. These five features cooperate together to form the partial discharge patterns that were fed to multilayer ANN trained using the back propagation algorithm. The same features and the ANN had been extended for the recognition of discharge sources of different types such as cavities and electric trees within the insulation system as well as the recognition of changes in the partial discharge shapes associated with the deterioration of

ageing effects within the defects undergoing discharge [Mazroua *et al*, 1995]. A new feature given by the product of the actual test voltage and the apparent charge as well as the width of the pulse were replaced by the multiplication of the pulse width and the apparent charge which resulted in six features. These were used to learn three different ANNs; the multilayer perceptron, nearest neighbour classifier and the linear vector quantization [Mazroua *et al*, 1994]. It was found that the recognition capabilities of the three ANNs were comparatively equivalent.

The back propagation feed forward ANN had also been used to discriminate the partial discharge of three kinds of electrode systems which resulted in 24 different classes [Okamoto *et al*, 1995].

In most of the above work concerning the application of ANN to partial discharge patterns recognition, multilayer perceptron with error back propagation as the learning algorithm has been used. This has provided a slightly better recognition rate when used for partial discharge recognition.

The main objective of this chapter has been to study the behaviour of a multilayer perceptron model of artificial neural network (ANN) using back propagation algorithm for partial discharge sources classification. The features of the four texture analysis algorithms as described in chapter 2 have been used to classify the six partial discharge sources as described in chapter-3. Detailed studies have been carried out to determine the features in each texture analysis method which give maximum classification accuracy as well as determining the optimum number of features which can be used for partial discharge classification.

6.2 Theoretical Background

In this work, a multilayer feed forward neural network has been used to classify different sources of partial discharge. There are certain aspects to be taken into account before one can obtain an ANN system that is able to cope with the partial discharge patterns recognition task. Essentially there are two stages to be followed: features extraction stage and classification stage. The first stage, which already has been done by using the texture analysis algorithms, is a step in which an m-dimensional pattern vector is mapped into a reduced n-dimensional feature vector, where n is equal to the number of extracted features. The feature vector are then applied to the ANN to perform the classification stage. The main idea behind the classification is to define a boundary surface that divides the features space into a number of disjoint regions that represent the different classes.

6.2.1 Basic architecture

The ANN mainly consists of processing elements and weighted connections [Haykin, 1994]. Fig. 6.1 shows an architectural graph of a multilayer feed forward neural network with single hidden layer. The neurons or the processing elements in the ANNs are arranged into three layers: input layer, output layer and hidden layer(s). The network shown here is fully connected in the sense that every node in each layer of the network is connected to every other node in the adjacent forward layer. The first layer is the input layer which receives respective elements of the input vector. Each neuron in the ANN collects the values from all of its input connections, perform a predefined mathematical operation (typically a dot product followed by a processing element function), and produce a single output value. The set of output signals of the neurons in the output layer of the network

constitutes the overall response of the network to the activation pattern supplied by the source node in the input layer.

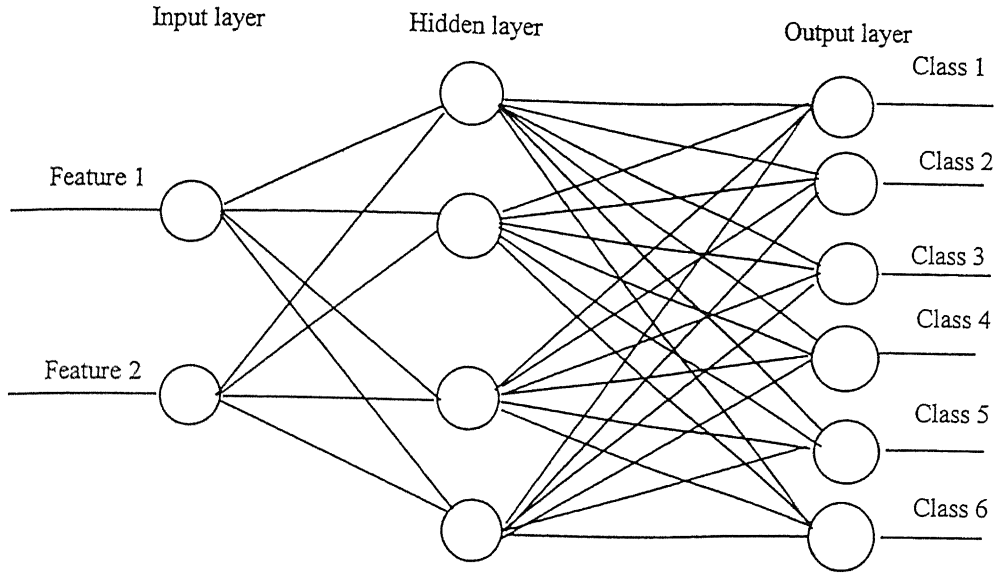


Fig. 6.1: A typical multi-layer perceptron model

6.2.2 The feed forward concept

In the feed forward neural network, signals can only propagate from the input layer to the output layer via one or more hidden layers. Only the nodes in the hidden layers and the output layer perform the activation function. Since the nodes in the input layer simply pass on the signal from the external source to the hidden layer, for each neuron in the input layer, the neuron output are given by:

$$O_i = N_i \tag{6.1}$$

where N_i is the input to the neuron i and O_i the output of the neuron i. For each neuron in the output layer, the inputs are given by:

$$n_k = \sum_{j=1}^{N_j} W_{kj} O_j \qquad k = 1, 2, \ldots\ldots\ldots\ldots, N_k \qquad (6.2)$$

Where, W_{kj} is the connection weight between hidden neuron j and output neuron k. N_j and N_k are the number of neurons in the hidden layer and output layer, respectively. The initial values for the connection weights are pre-set to small random values within a predefined range. The processing element activation functions map the processing element inputs within a prespecified range. Although the possible number of processing element activation functions is infinite, in this chapter the most common sigmoid function has been used. The neuron outputs are given by:

$$O_k = \frac{1}{1 + e^{-(n_k + \theta_k)}} = f(n_k, \theta_k) \qquad (6.3)$$

where, θ_k is an externally applied threshold that has the effect of decreasing or increasing the net input of the activation function. For the neurons in the hidden layer, the input and the output are given by relations similar to that given in equations (6.2) and (6.3), respectively.

6.2.3 The back propagation concept

The neural network development involves training and testing phases. In the training phase, the weights of the ANN are adjusted to map the input of the system to its output. In the testing phase, the ANN should predict the correct system output for a given set of inputs, even if these were not used in training. An untrained or poorly trained network will give erroneous output. Therefore, as a measure of how a network is functioning during the training, the output at the last layer has to be evaluated. There is an algorithm based on the

minimisation of the error function on each pattern P by using the steepest descent method. The sum of the squared errors E_p which is the error function for each pattern is given by:

$$E_p = \frac{1}{2}\sum_{k=1}^{N_k}(t_{pk} - O_{pk})^2 \tag{6.4}$$

where t_{pk} is the target output for output neuron k and O_{pk} is the calculated output for the output neuron k. The overall measure of the error for all the input-output patterns is given by:

$$E = \sum_{P=1}^{NP} E_p \tag{6.5}$$

where, NP is the number of input-output patterns in the training set. The learning procedure used is the back propagation algorithm [Lippmann, 1987]. When an input pattern P with the target output vector t_p is presented, the connection weights between hidden and output layers are updated by using the following equations:

$$W_{kj}(p) = W_{kj}(p-1) + \eta\delta_{pk}O_{pj} + \alpha\Delta W_{kj}(p-1) \tag{6.6}$$

$$\delta_{pk} = (t_{pk} - O_{pk})O_{pk}(1 - O_{pk}) \tag{6.7}$$

where, η is the learning rate, and α is the momentum constant. The learning rate controls the width of the steps that must be moved from the current position of the error surface down to the global minimum. A small η results in a very slow convergence whereas a large value of η will obviously lead to a faster convergence but it may also be accompanied by oscillation [Lippmann, 1987]. Momentum is a kind of memory that incorporates the weight change of the previous step and in this way slows down useless oscillations. Similarly the correction of weight between the input layer neuron i and hidden layer neuron j can be updated by using the following equations:

$$W_{ji}(p) = W_{ji}(p-1) + \eta \delta_{pj} O_{pj} + \alpha \Delta W_{ji}(p-1) \tag{6.8}$$

$$\delta_{pj} = O_{pj}(1 - O_{pj}) \sum_{k=1}^{N_k} \delta_{pk} W_{kj} \tag{6.9}$$

The effect of threshold is represented by adding a new input signal fixed at unity and a weight equal to the threshold value. Therefore the thresholds are found exactly in the same manner as the connections weights.

6.2.4 Modified back propagation method

The back propagation algorithm with sigmoid function, as described above, encounters the following difficulty. When the actual output value approaches either extreme value, the factor $(O_{pi} - 1) \times O_{pi}$ in equation (6.7) makes the error signal very small. This implies that the output unit can be maximally wrong without producing a strong error signal with which the weights could be significantly adjusted. For instant, this occurs when some of the output values are pushed towards the wrong extreme value by competition in the network, thereby not increasing their error signal but instead decreasing it. In order to overcome this problem, instead of minimising the square of the differences between the actual and target values summed over the output units, the minimisation of the following modified error function has been used [Van Ooyen et al, 1992].

$$E_p = -\sum_{k=1}^{N_k} [t_{pk} \ln O_{pk} + (1 - t_{p_i}) \ln(1 - O_{p_i})] \tag{6.10}$$

The above function resulted in

$$\delta_{pk} = (t_{pk} - O_{pk}) \tag{6.11}$$

Thus, the error signal propagating back from each output unit is now directly proportional

to the difference between the target value and the actual value. The modified back propagation method based on equation (6.10) has been used in the present work. The training data used in this chapter were obtained from the experimental work discussed in chapter 3. Total 180 patterns were generated to describe the six classes i.e. for each class 30 patterns were obtained. Fifty percent of these patterns were used for training, while all the patterns were used for testing the ANN.

6.3 Experimental Design

The main aim of the this chapter was to apply ANN to investigate the discriminating power of the texture features in classifying different partial discharge sources. Based on the relative discriminating power, the optimum number of features, for the desired classification accuracy, could be selected. This has been achieved in two steps. The first step utilised each feature individually for partial discharge classification. From this step, one can determine the maximum classification accuracy, while using only one feature for the classification, as well as the best features for each technique. In the second step, all combinations of two features at a time were utilised and the best combination was established based on the maximum classification accuracy. Number of the extracted features corresponding to the texture analysis techniques were 28, 22, 20 and 14 for SGLDM, GLDHM, GLRLM, and PSM, respectively. This involved 84 features for the four techniques when used individually plus 890 combinations of two features (378 for SGLDM, 231 for GLDHM, 190 for GLRLM and 91 for PSM).

It is well known that the performance of the ANN is controlled by different parameters like the learning rate, the momentum, number of neurons in the hidden layer and the number of

training iterations. All these parameters depend upon the input data. Therefore, to get the best classification accuracy, all these parameters have to be adjusted for the given application. The details of the selection of these parameters in the present work are given below.

Since the ANN in the present work has been used as classifier for partial discharge sources considering each feature individually, which have different discriminating power, or combination of different features, the learning rate has not been fixed. A self adaptation method [Battiti, 1989] based on the gradient descent method has been used to adjust the learning rate value. One starts with a learning rate η, in this study 0.3 was selected as a starting value. If the error function E to be minimised decreases, one updates $\eta = \rho\eta$ where $\rho = 1.1$ in each iteration until E increases. One then reduces $\eta = \sigma\eta$ with $\sigma = 0.5$ and iterates until a step when decreasing E is found.

The number of neurons in the input layer was chosen same as the number of features used for the ANN training, in this case either one or two. Six neurons were used in the output layer to represent the six possible partial discharge sources. For each partial discharge source, only one neuron gives high output (one) and all the other five neurons give low output (zero). In the hidden layer, four neurons were used for both the cases of one or two input features. Number of neurons in the hidden layer was fixed based on hit and trial and was taken as a number resulting in the minimum error and the minimum number of training iterations. Also, it was experimentally found from hit and trial that the momentum constant value of 0.5 gives satisfactory results.

6.3.1 Effect of number of training iterations

One training iteration is defined as learning all training patterns once. To investigate the effect of number of training iterations, the mean of the squared output error has been obtained for 100 iterations for all the features of the four techniques. Figs. 6.2a to 6.2c show the relation between the iteration number and the mean square error for the 22 features of GLDHM for the case of single feature input. This figure shows that:

(i) There are two classes of features; one giving almost constant mean square error for any number of learning iterations, the second for which the mean square error decreases to a certain value and then remain constant.

(ii) There is no considerable change in the mean square error value for all the features after the first 50 iterations except for a few for which error almost saturates in 100 iterations.

To investigate more deeply the effect of increasing the number of training iterations on the classification accuracy of the ANN, the features of every techniques were ranked according to their relative classification accuracy. Then the best five features for each technique, which have the maximum classification accuracy after 100 learning iterations, were retrained by increasing the maximum number of iterations to 1000. Similarly, the ANN which was trained using combination of two features, was also retrained. While it was trained originally until the mean square error reduced to 10% or for maximum of 300 iterations, the retraining of the ANN was carried out with the best five combinations of two features until the mean square error reduced to 1% or maximum 3000 iterations. The testing of the ANN were performed to establish their classification accuracy. The results of classification accuracy for the features used individually are shown in table 6.1, and the

results with combinations of two features are shown in table 6.2.

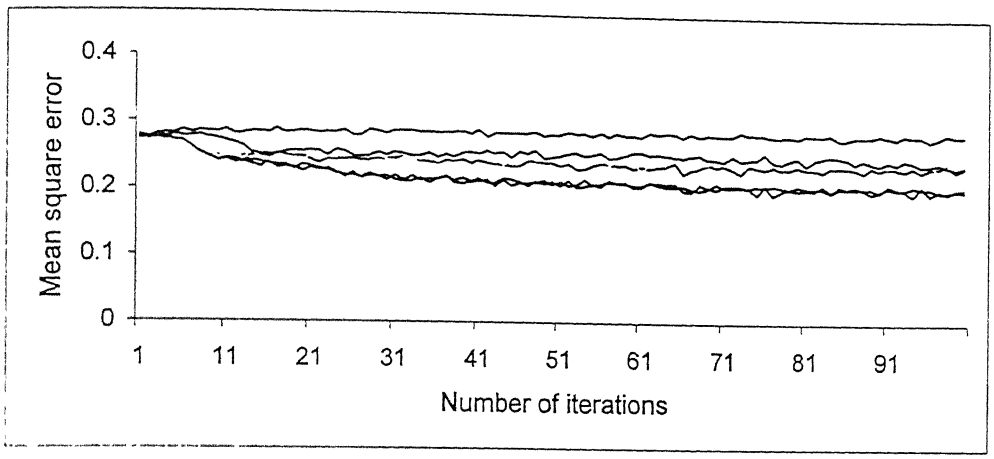


Fig. 6.2a: Plot of mean square error in the learning process for features from 1 to 7 with GLDHM

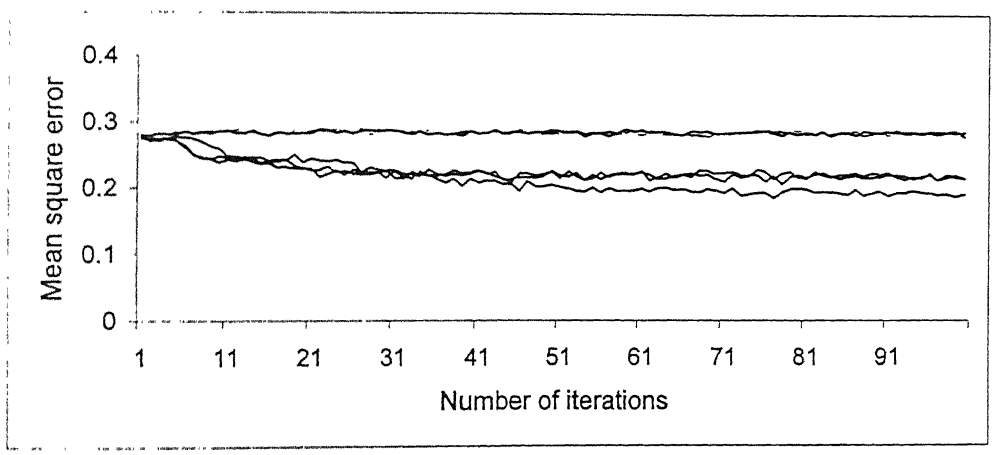


Fig. 6.2b: Plot of mean square error in the learning process for features from 8 to 14 with GLDHM

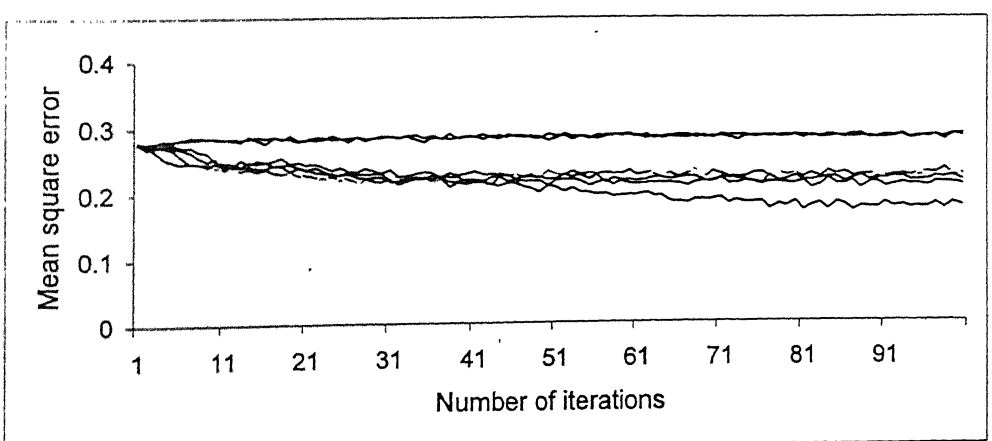


Fig. 6.2c: Plot of mean square error in the learning process for features from 15 to 22 with GLDHM

With increasing the number of iterations with individual features from 100 to 1000, training time increased to almost 10 times. However no substantial increase in classification accuracy was obtained. In some of the cases, the classification accuracy has even reduced. With increasing the number of iterations for the training with the combinations of two features from 300 to 3000 iterations, the variation in the classification accuracy was even smaller compared to the case of one feature. The variation was within 5% except for one case which was around 14% for the combination of features 3 & 13 of GLRLM.

Table 6.1: Effect of increasing the number of training iterations of ANN with one feature input

	Feature number	Classification accuracy		Mean Square error	
		100 iterations	1000 iterations	100 iterations	1000 iterations
GLDHM	17	79.44	76.66	17.8	15.9
	21	57.77	54.44	21.1	22.2
	14	57.77	61.11	21.6	21.0
	15	51.66	48.88	22.6	21.4
	12	51.55	57.77	21.5	20.4

	Feature number	Classification accuracy		Mean square error	
		100 iterations	1000 iterations	100 iterations	1000 iterations
SGLDM	17	55	57.22	21.7	20.7
	24	54.44	66.11	21.8	19.0
	20	50	70	21.4	18.6
	18	46.66	48.88	22.2	21.3
	7	43.88	65	18.6	19.6

	Feature number	Classification accuracy		Mean square error	
		100 iterations	1000 iterations	100 iterations	1000 iterations
GLRLM	10	66.66	66.66	18.7	18.5
	3	55	57.77	20.4	18.6
	20	50	33.33	22.16	22.0
	11	50	58.88	21.4	21.5
	13	48.33	47.22	22.4	22.9

	Feature number	Classification accuracy		Mean square error	
		100 iterations	1000 iterations	100 iterations	1000 iterations
PSM	10	32.22	42.22	24.9	24.1
	14	28.33	28.88	25.7	24.7
	5	27.77	2.77	27.0	26.9
	7	26.11	27.77	26.2	27.1
	13	22.77	25.55	24.0	24.5

Table 6.2 : Effect of increasing the number of training iterations of ANN with combination of two features input

	Feature	Maximum iterations =300			Maximum iterations =3000		
		Classification accuracy	Mean square error	Actual iterations	Classification accuracy	Mean square error	Actual iterations
GLDHM	9&20	94.44	9.61	1OO	95.55	0.95	1677
	10&21	93.33	11.22	300	90.55	0.99	1795
	3&20	92.77	9.29	145	95	4.49	3000
	7&17	92.22	9.81	111	96.11	0.97	2383
	4&17	91.66	9.47	22	96.66	3.45	3000

	Feature	Maximum iterations =300			Maximum iterations =3000		
		Classification accuracy	Mean square error	Actual iterations	Classification accuracy	Mean square error	Actual iterations
SGLDM	10&24	96.11	9.93	63	95.55	0.95	1677
	10&27	95.55	9.59	267	96.11	0.83	1275
	6&17	93.88	9.90	153	96.11	6.10	3000
	13&20	93.33	9.06	107	97.22	0.91	660
	10&17	93.33	9.46	54	93.33	5.85	3000

	Feature	Maximum iterations =300			Maximum iterations =3000		
		Classification accuracy	Mean square error	actual iterations	Classification accuracy	Mean square error	Actual iterations
GLRLM	10&13	80.55	12.73	300	78.33	13.2	3000
	3&20	79.44	9.80	248	83.88	9.27	3000
	3&13	75.55	16.83	100	89.44	11.4	3000
	3&5	67.77	17.86	100	69.44	16.7	3000
	10&17	66.66	18.76	100	66.44	18.2	3000

	Feature	Maximum iterations =300			Maximum iterations =3000		
		Classification accuracy	Mean square error	Actual iterations	Classification accuracy	Mean square error	Actual iterations
PSM	5&13	52.77	21.99	100	53.88	20.20	3000
	6&12	49.44	23.57	100	51.66	22.72	3000
	3&12	48.88	23.93	100	60.00	20.96	3000
	7&12	45.00	23.83	100	48.33	22.53	3000
	12&13	43.88	24.12	100	41.66	24.23	3000

Since the features, which make the training process faster, were considered as the best features, based on the above results, the training was performed for the mean square error

to be less than 10% or maximum number of 100 iterations in the case of individual features and 300 iterations in the case of two features combination.

6.3.2 Normalisation of inputs and weight initialisation

Scaling of the input-output data has a significant influence on the convergence property and also on the accuracy of the learning process. It is obvious from the sigmoidal activation function that the range of output of the network must be within 0 to 1. Moreover the input variables should be kept small in order to avoid saturation caused by the sigmoidal function. To normalise the input data, the maximum value of the input vector components were determined as follows[El-Makkawy, 1998]:

$$n_{i,max} = \max (n_i(p)) \qquad p=1,\ldots\ldots\ldots NP \qquad i=1,\ldots\ldots\ldots\ldots N_i \tag{6.10}$$

where NP is the number of patterns in the training set and N_i is the number of neurons in the input layer. The input data were normalised by this maximum value as the following:

$$n_{i,nor}(p) = n_i(p) / n_{i,max} \quad p=1,\ldots\ldots\ldots\ldots NP \quad i=1,\ldots\ldots\ldots\ldots\ldots N_i \tag{6.11}$$

After normalisation, the input variables range is within 0 to 1. The initial weights of the ANN were randomly initialised in a range of -0.01 to 0.01.

6.4 Results and Discussion

The ANN utilised in this study had only one hidden layer with 4 neurons. The input patterns is a vector of length one or two depending on the number of features used for the training of the ANN. For both the cases of one or two inputs, the ANN had six output neurons corresponding to six different partial discharge sources. It is expected that after the ANN has completed its learning phase, even for the unknown test patterns, only one of the

six outputs is high or equal to one while the other five outputs are equal to zero. By using each feature individually for the training of the ANN, the classification accuracy of all the features obtained from the ANN testing corresponding to the four techniques are shown in Figs. 6.3 to 6.6. The classification accuracy has been calculated as the percentage of patterns out of the original 180 patterns which were classified correctly during the phase of the ANN testing.

From Fig.6.3, it is clear that with SGLDM, the ANN was able to learn better by using the features which describe the horizontal direction (features 1-7 and 15-21) than the vertical direction (features 8-14 and 22-28) especially in the negative half cycle. The correlation between the partial discharge pulses in the horizontal direction (features 7, 21) has a better classification accuracy compared with the same feature when used for the vertical direction (features 14, 21). It was also observed that the ANN has learnt better the features which describe the negative half cycle (features 15-22) in comparison with positive half cycle (features 1-14).

With the GLDHM, as shown in Fig. 6.4, the maximum classification accuracy was obtained by feature 17. This feature is related to the horizontal direction of the negative half cycle. In the positive half cycle, the maximum classification accuracy was achieved by feature number 1, which is also related to the horizontal direction. In this case also, it was noticed that the features which describe negative half cycle (features 12-22) are more effective in the classification process compared to the positive half cycle (features 1-11). In the negative half cycle, the features belonging to the horizontal direction are more effective in partial discharge classification compared with the vertical direction.

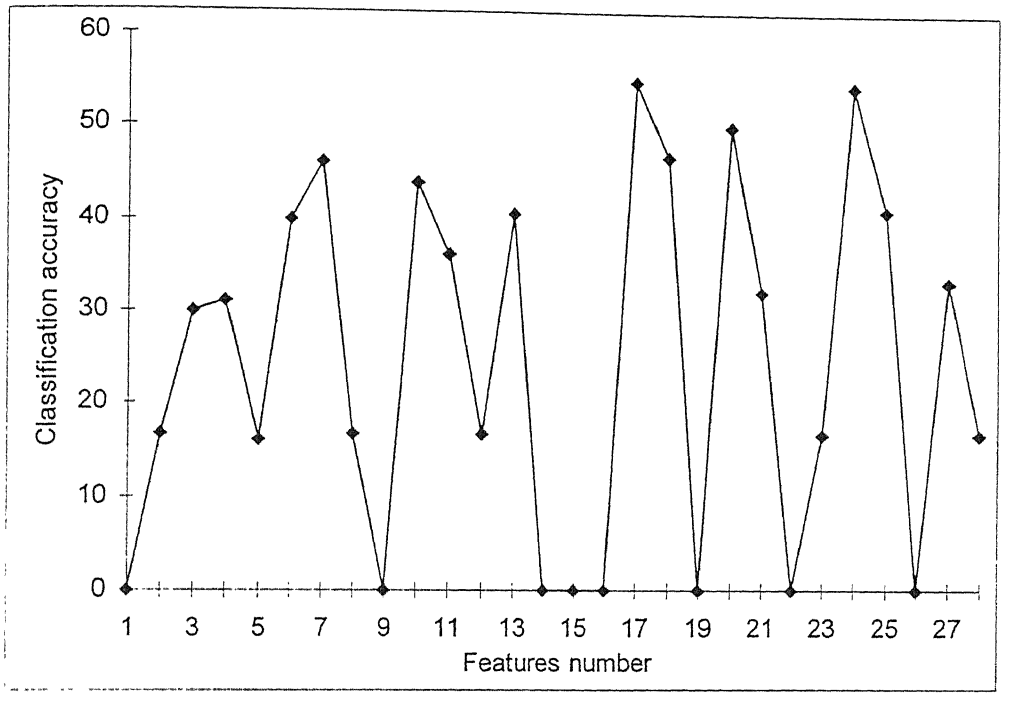


Fig. 6.3: Classification accuracy of SGLDM features

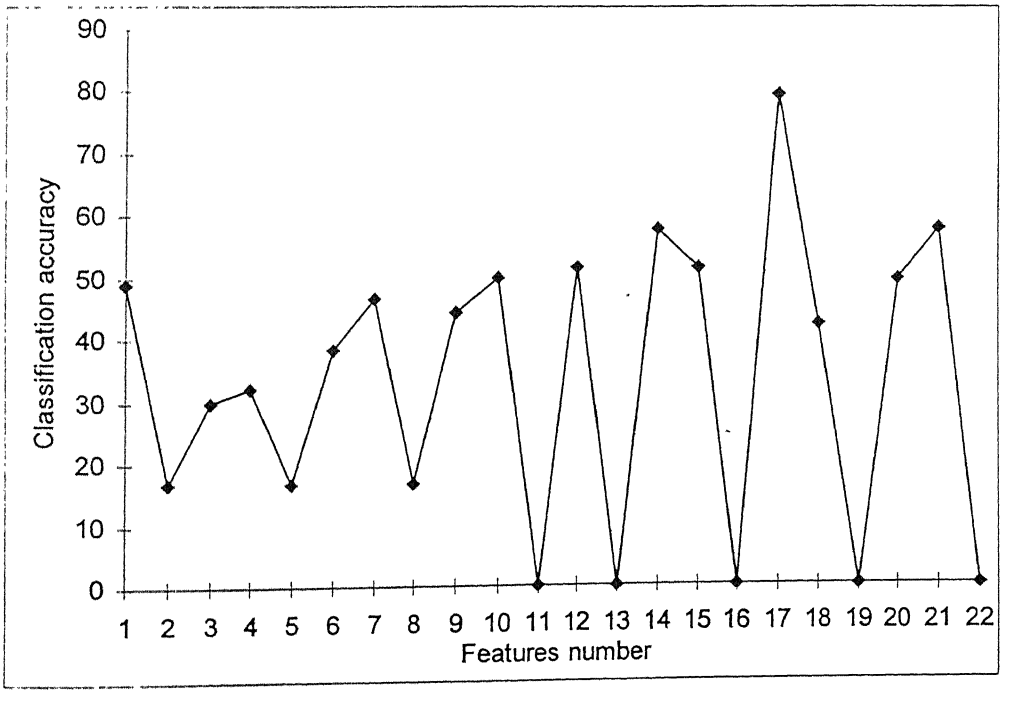


Fig. 6.4: Classification accuracy of GLDHM features

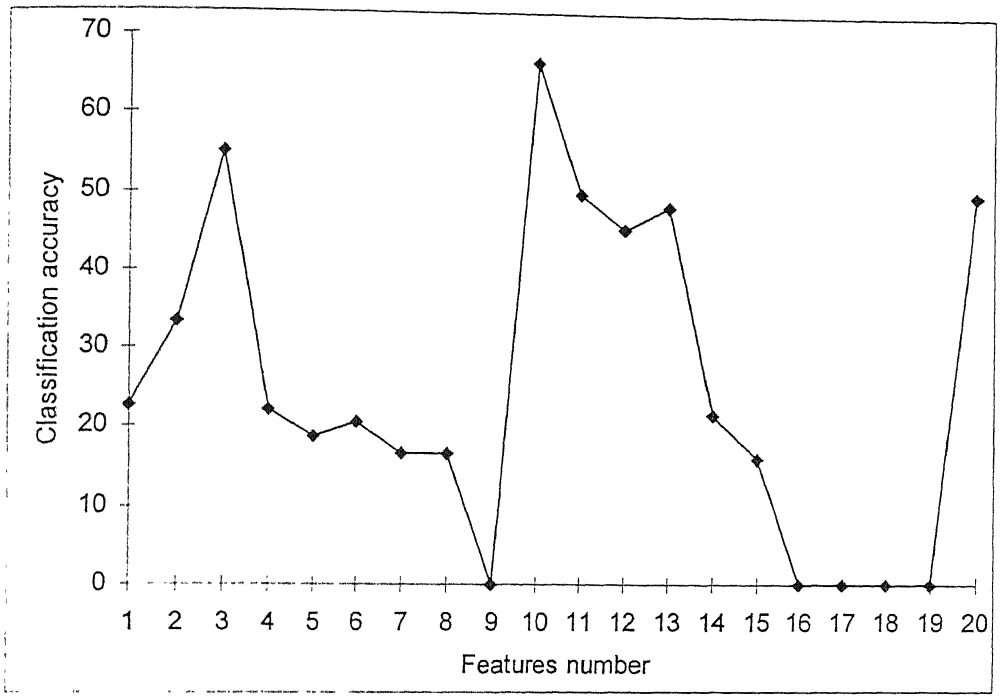


Fig. 6.5: Classification accuracy of GLRLM features

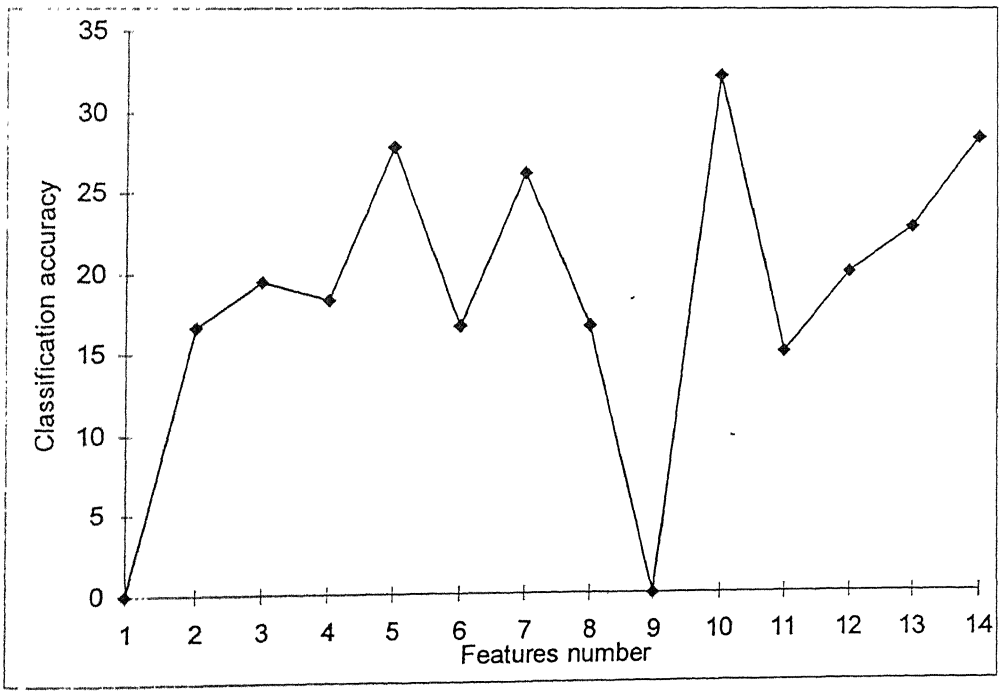


Fig. 6.6: Classification accuracy of PSM features

With GLRLM, as shown in Fig. 6.5, the vertical direction of the positive half cycle (features 6-10) is more effective in the classification process compared with the vertical direction of the negative half cycle (features 16-20). However, the horizontal direction of the negative half cycle is much better than the horizontal direction of the positive half cycle. The maximum classification accuracy was achieved by feature number 10. In the same cycle, the features which describe the horizontal direction (features 1-5 and 11-15) are more effective in classification except features 5 and 15 which have classification accuracy less than the features 10 and 20, respectively.

With PSM, as shown in Fig. 6.6, the classification accuracy was very poor compared to the other techniques. The maximum classification accuracy was 32.22% for the feature number 10.

Comparing the classification accuracy achieved by using the features of the texture analysis algorithms with the ANN with the results of the minimum distance classifier, it was noticed that:

(i) The classification accuracy of the features using the ANN are generally smaller than the classification accuracy of the same features using the minimum distance classifier.

(ii) The ANN could not be trained by using certain features which resulted in zero classification accuracy such as, for example, features 1,9 for SGLDM, features 11,13 for GLDHM, features 9,16 for GLRLM and features 1,9 for PSM .

The poor classification accuracy of the ANN might have resulted from the overlapping

between the partial discharge sources in the feature space. To investigate the effect of overlapping between the classes on the classification accuracy, the scatter of the six sources of partial discharge by using the best and worst features (which having the maximum and minimum classification accuracy) for GLDHM are shown in Figs. 6.7 & 6.8. The horizontal axis is the number of patterns from 1 to 180. Patterns 1 to 30, 31 to 60, 61 to 90, 91 to 120, 121 to 150 and 151 to 180 belong to classes 1, 2, 3, 4, 5 and 6, respectively. The vertical axis represents the values of corresponding feature. It is clear that the classes are more separate with feature 17 since there is no overlapping between classes 1, 3, 4, 5 and 6. However, they become closer with feature 11, since there is clear overlapping between the classes 1, 3, 4, 6. Also, there is overlapping between class 2 and 5. For PSM, which gives the worst result, the scatter of feature 4 is shown in Fig. 6. 9 which shows that all the classes are overlapped.

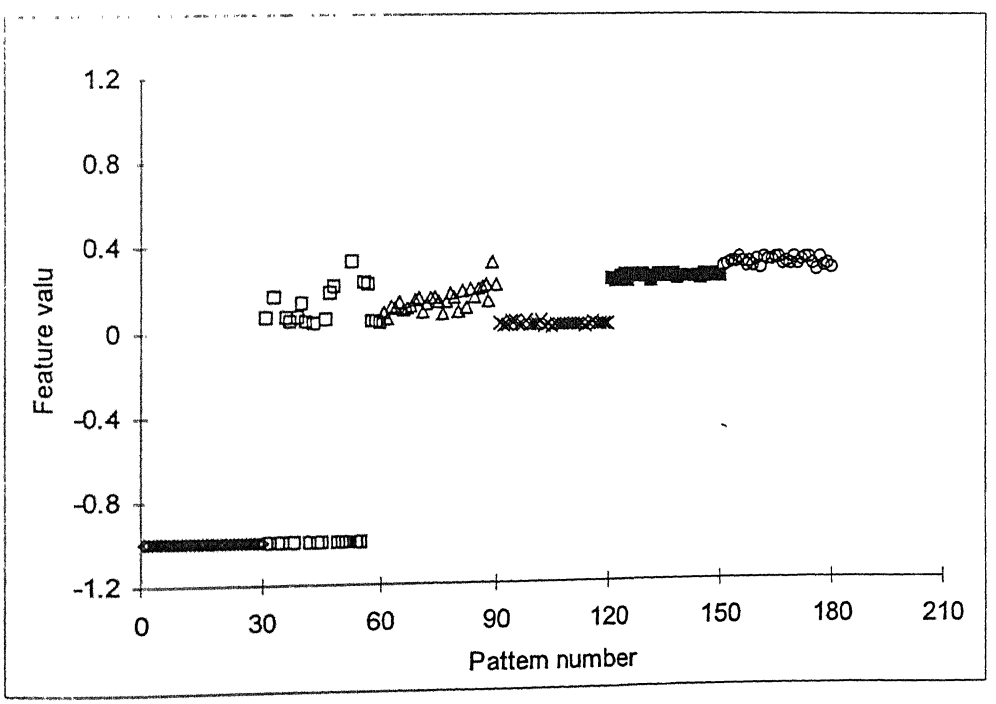


Fig. 6.7: Distribution of partial discharge patterns using feature 17 of GLDHM

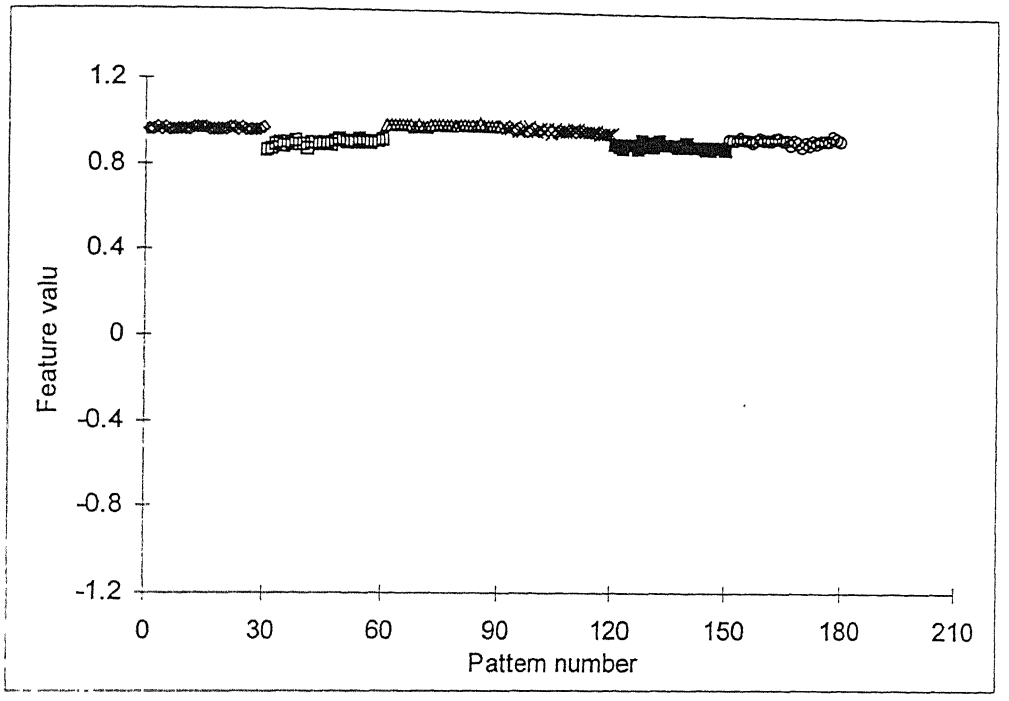


Fig. 6.8: Distribution of partial discharge patterns using feature 11 of GLDHM

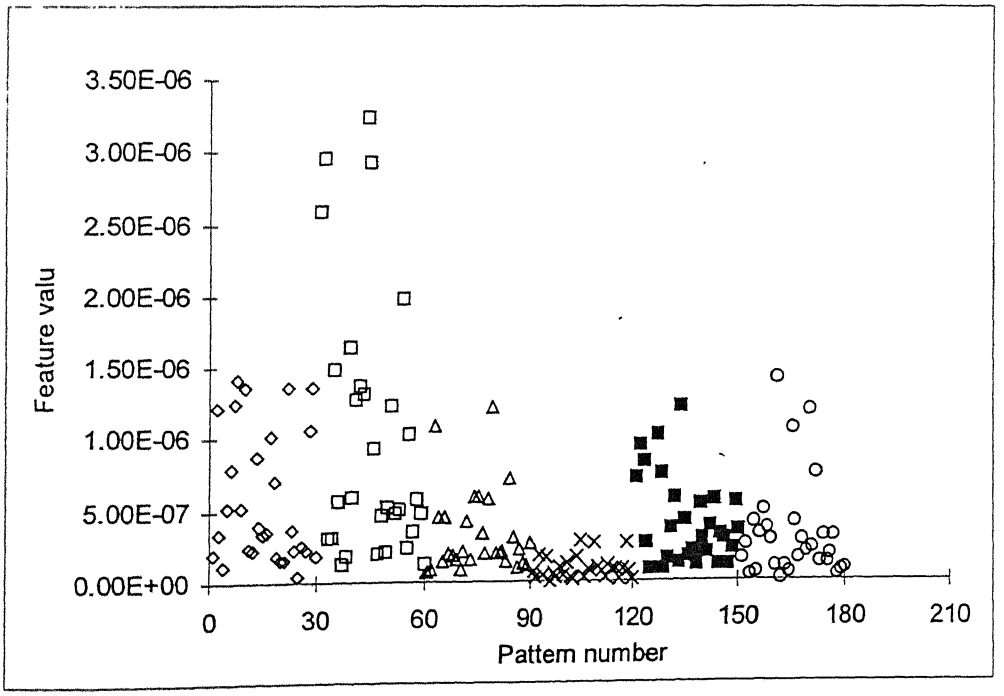


Fig. 6.9: Distribution of partial discharge patterns using feature 4 of PSM

6.5 Effect of Increasing the Number of Features Used for Classification

By studying the classification accuracy using only one feature at a time, it is observed that the maximum classification accuracy achieved is around 80% in the case of GLDHM. Hence one feature is not sufficient to achieve a good classification accuracy between the partial discharge sources. Therefore, in the next step, the effect of using two features at the same time for the training of the ANN were studied. The classification accuracy of SGLDM, GLDHM, GLRLM and PSM by using all the possible combinations of two features for each technique were investigated. The most effective twenty combinations of features for each technique are shown in Figs. 6.10 to 6.13.

From Fig. 6.10, one can notice that all the combinations of features, which gave the best classification accuracy, generally had one feature belonging to the positive half cycle and the second feature to the negative half cycle except the combination (7,11) which belongs to the positive half cycle only. It is further noticed that these combinations are equally divided into four groups. The first group is related only to the horizontal direction of the positive and negative half cycles e.g. (3,17), (3,20), (6,17). The second group is related only to the vertical direction of the positive and negative half cycles e.g. (10,24), (10,27), (13,27). The third group is combination from the horizontal direction of the positive half cycle with the vertical direction of the negative half cycle e.g. (3,27), (4,25), (6,24). And the fourth is a combination from the vertical direction of the positive half cycle and the horizontal direction of the negative half cycle e.g. (10,20), (11,18), (13,17).

From Fig. 6.11, it is clear that all the combinations of features which have the maximum classification accuracy are related to the positive and negative half cycles except the combination (4,6) in which both features belong to the positive half cycle only. All the

combinations were distributed between horizontal-horizontal e.g. (1,17), (4,17), (6,17), vertical-vertical e.g. (7,18), (9,20), (10,21), and mixed from both directions e.g. (1,18), (3,20), (7,17), (10,12). By using GLRLM, out of the best 20 combinations, 8 combinations were related only to the positive half cycle, as shown in Fig. 6.12. From Fig. 6.12, it is interesting to note that the feature number 10, which was the best when used individually, is common in most of the combinations e.g. (3,10), (4,10), (5,10), (6,10).

With PSM, the situation is slightly different where the combinations were related to positive and negative half cycles e.g. (3,12), (3,13) as well as other combinations related to the negative half cycle only e.g. (10,11), (10,12). By using combinations of two features, PSM classification accuracy has improved but still shows a poor classification accuracy compared to the other techniques. The maximum classification accuracy for the combinations of two features extracted from the four techniques were 96.11, 94.44, 80.55 and 52.77 for SGLDM, GLDHM, GLRLM and PSM, respectively. This means that SGLDM and GLDHM have good discriminating power compared to GLRLM and PSM.

To investigate the effect of the overlapping between the classes, the scatter distributions of the six classes of partial discharge sources, according to the best and the worst combination of two features for GLDHM, are shown in Figs. 6.14 & 6.15. It can be observed that by using the combination (9,20), which is the best combination, the classes 1, 2, 5, 6 are clearly separable. While using the combination (12,17), only classes 4 and 5 are clearly separable. The scatter of the same classes according to the combination (3,4), which belongs to PSM, is shown in Fig. 6.16. From Fig. 6.16, it is extremely difficult to recognise any class because of the high overlapping between the classes which resulted in the poor classification accuracy.

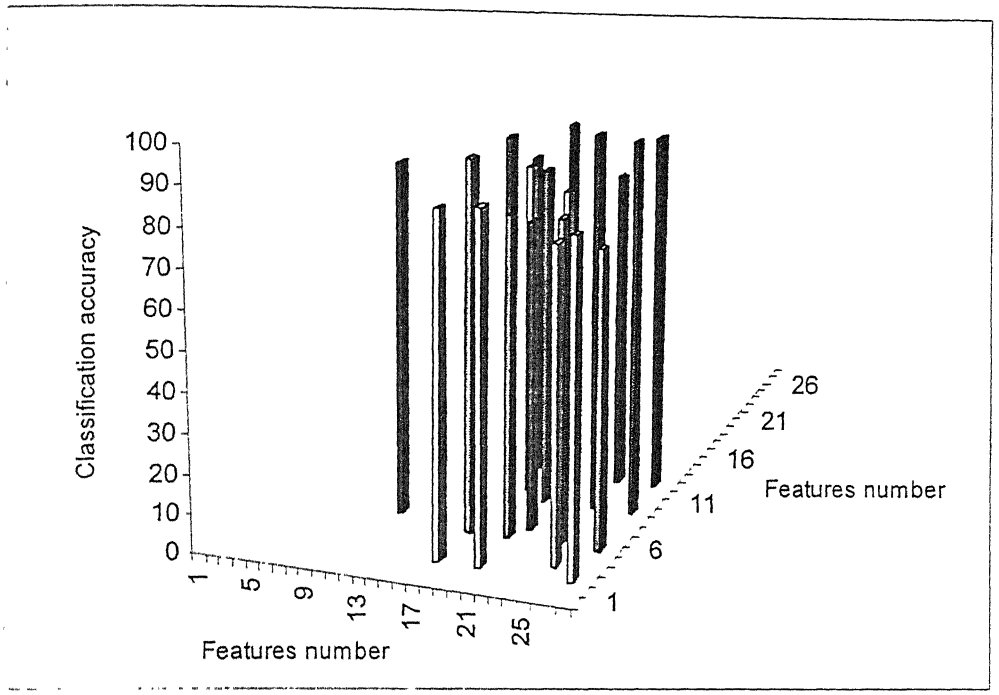


Fig. 6.10: Classification accuracy of SGLDM features combinations

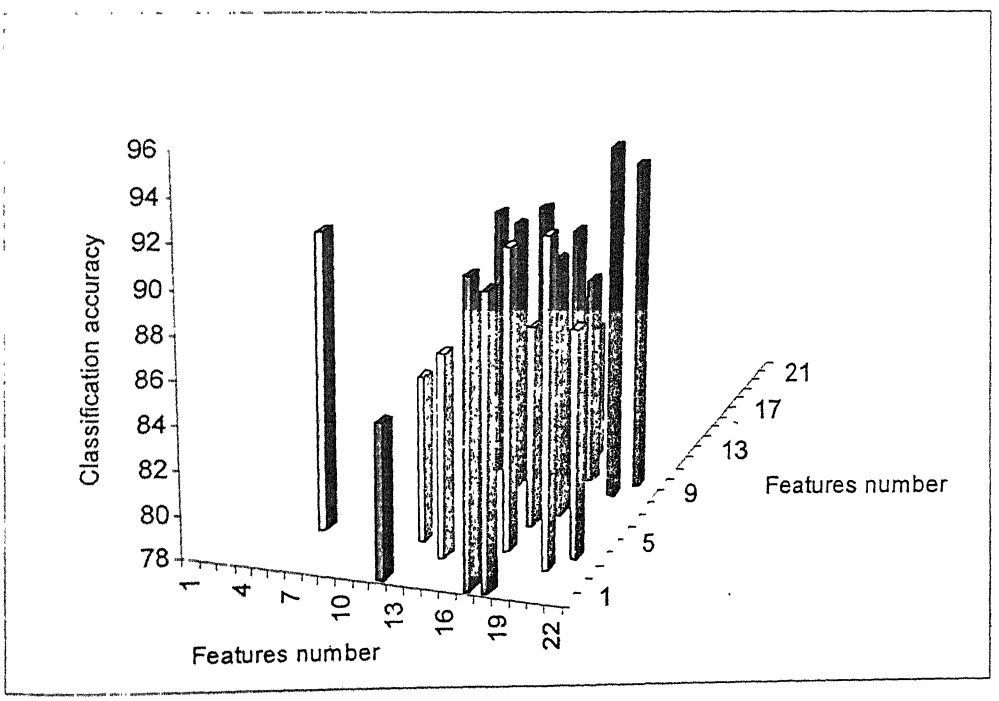


Fig. 6.11: Classification accuracy of GLDHM features combinations

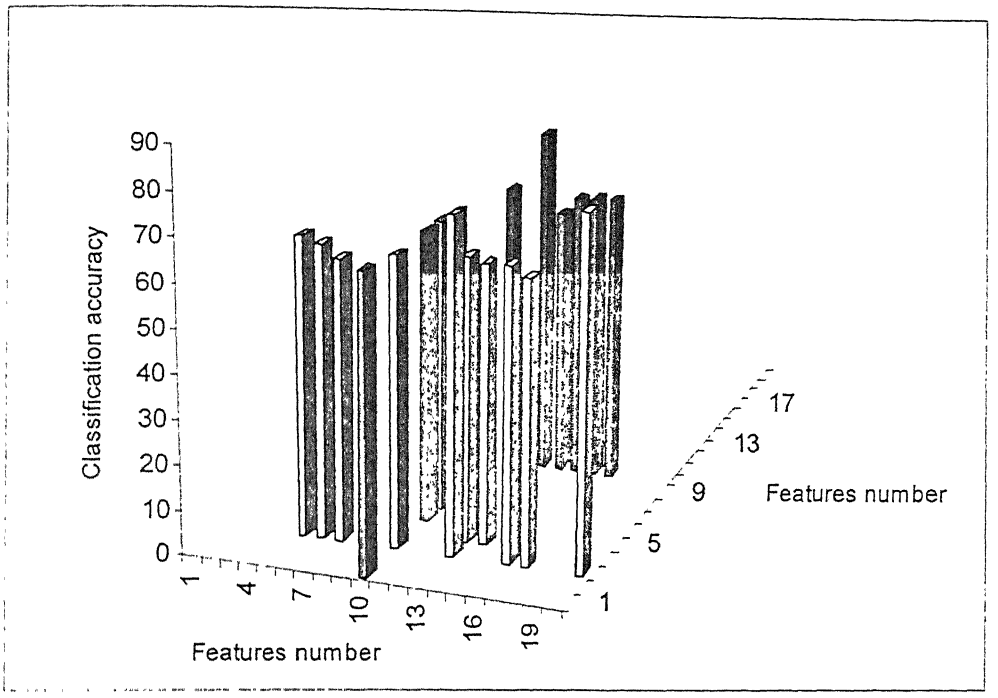


Fig. 6.12: Classification accuracy of GLRLM features combinations

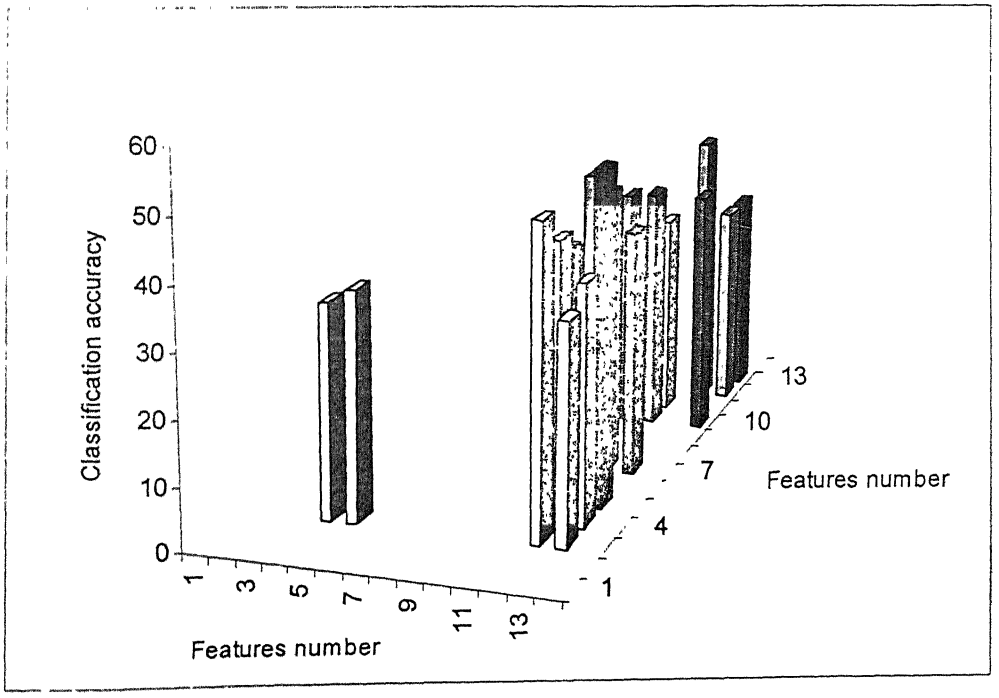


Fig. 6.13: Classification accuracy of PSM features combinations

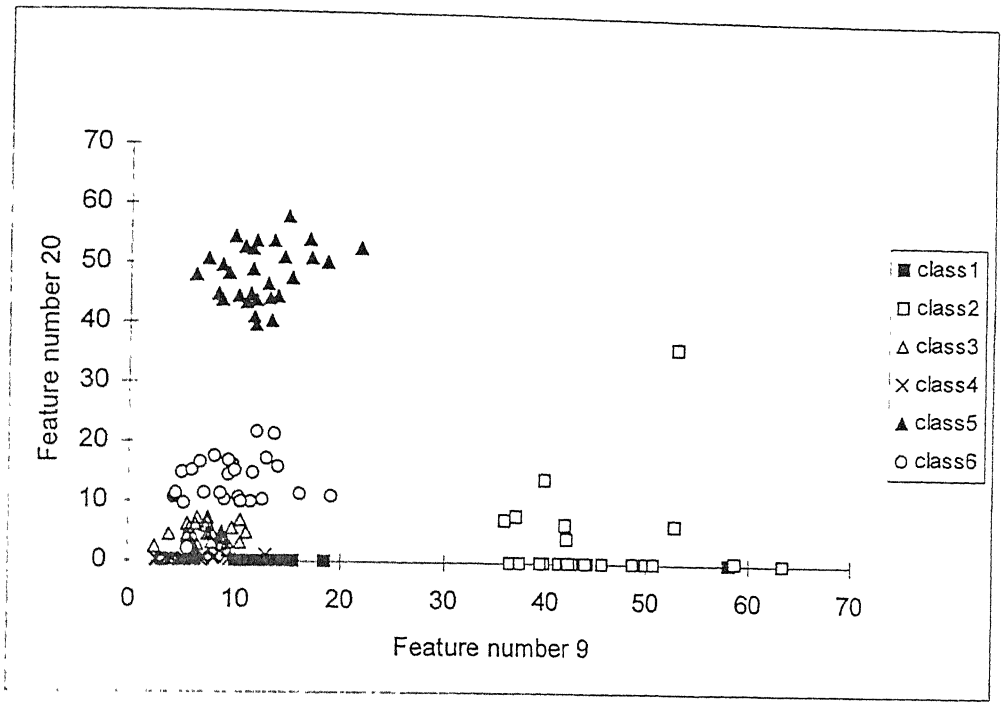


Fig. 6.14: Distribution of partial discharge patterns using features (9,20) with GLDHM

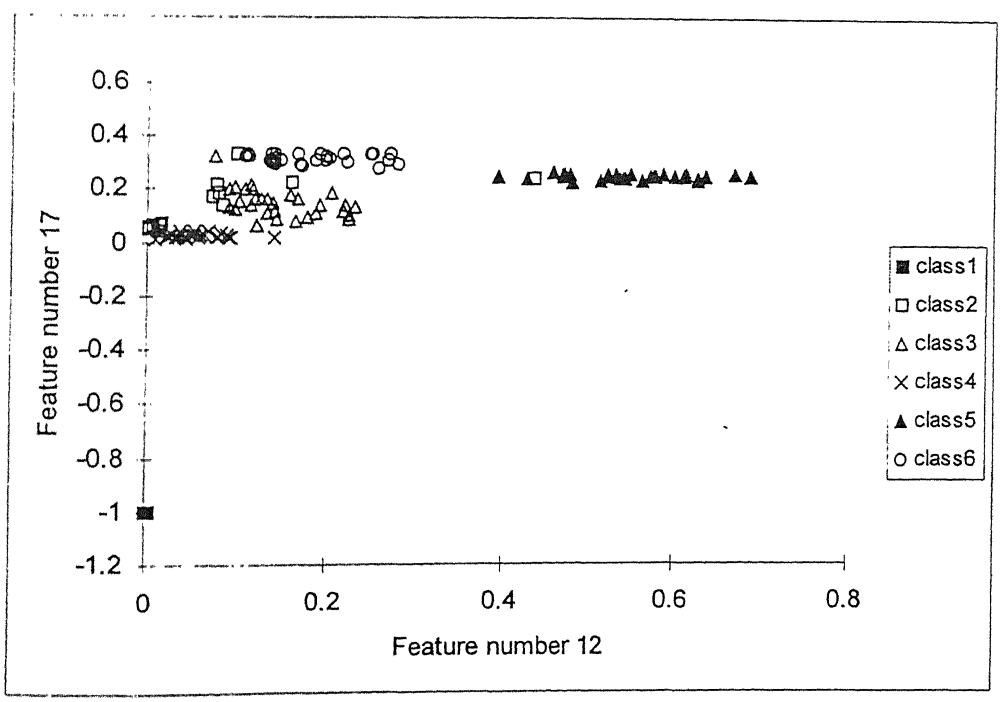


Fig. 6.15: Distribution of partial discharge patterns using features (12,17) with GLDHM

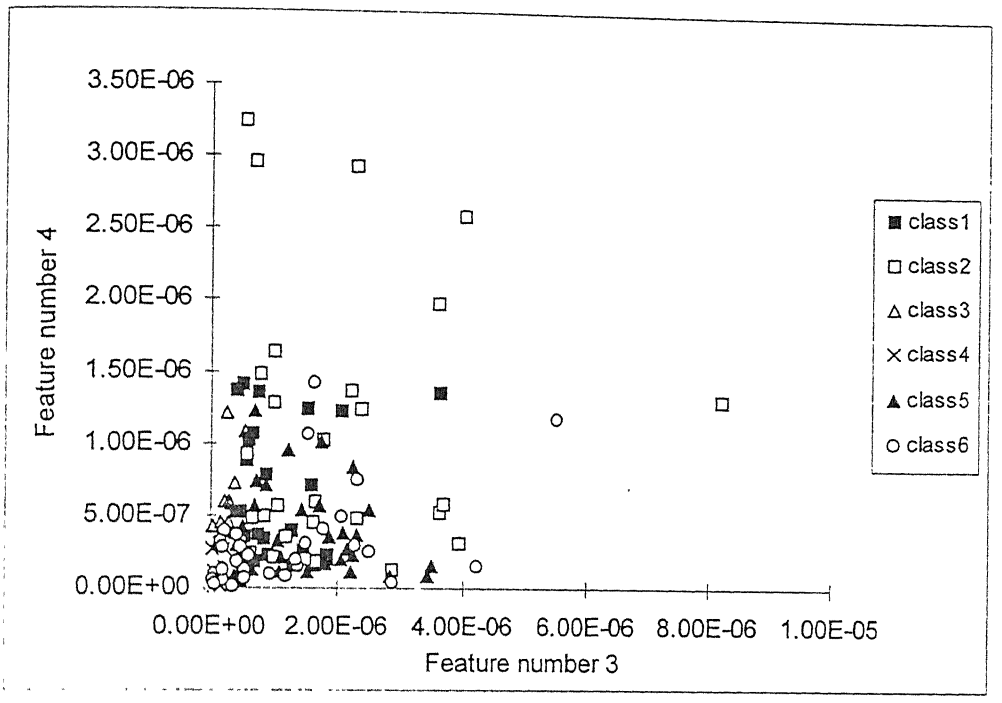


Fig. 6.16: Distribution of partial discharge patterns using features (3,4) with GLDHM

6.6 Conclusions

In his chapter a multilayer feed forward neural network, based on modified back propagation training algorithm, has been used for the classification of different partial discharge sources. The ANN was trained for the partial discharge sources representing glow corona, streamer corona, surface discharge, internal discharge, single protrusion, and multi protrusions. Based on the results reported in this chapter, following main conclusions can be drawn:

1- The Back propagation ANN with input features derived from the texture analysis techniques are capable of distinguishing between patterns of the six types of simple partial discharge sources with appropriate selection of feature combinations for its

training.

2- Since some of the partial discharge sources deviate from normal distribution, it was expected that the ANN will improve the classification accuracy. However, it is found that the classification accuracy slightly reduced as compared to the minimum distance classifier results reported in chapter 4. This may be due to the overlap between the attributes of different partial discharge sources.

3- Only two features are sufficient to obtain acceptable classification accuracy. More than two feature combinations will increase the computational time substantially.

4- Out of the four texture analysis techniques, the SGLDM and GLDHM provide almost the same classification accuracy. GLRLM has less discriminative power compared to the SGLDM and GLDHM. The PSM provides poor classification accuracy.

Chapter 7

Partial Discharge Patterns Classification for a Practical Case

7.1 Introduction

In the previous chapters, the discriminating power of the texture analysis algorithm features have been examined by using six standard partial discharge sources created in the laboratory with the help of different electrode systems. It was proved that the GLDHM, SGLDM and GLRLM are able to distinguish the different partial discharge sources by using only two features at a time. It was felt worth examining the discriminating power of these features using practical partial discharge sources. In this chapter, a cable test sample has been used for this purpose. Both internal and surface discharge are expected to be available in the cable sample. The classifier were used to distinguish between the internal discharge patterns and the internal discharge plus the surface discharge patterns.

7.2 Background

In addition to various obvious and by far the most usual reasons for failure of cables, such as mechanical damage to the insulation, ingress of moisture, thermal breakdown, and breakdown under transient voltage conditions, the partial discharge is also a basic reason for its failure[Bungay, 1990]. Cables are usually tested for the presence of the partial discharge activity, before installation, to detect any insulation defects that may have

occurred during the manufacturing process. Partial discharge tests are also used to detect the insulation deterioration in the normal service operating conditions. Partial discharge in a cable are caused by the breakdown of the gas contained within voids in the insulation. The voids may be either dielectric-bounded or at the interface between dielectric and conducting screens or conductor.

All cables are subjected to changes in load and therefore to temperature cycling during operation. The change due to the thermal expansion of both the conductors and the insulating material under the metal sheath, for example in paper insulating cables, produce small cavities (voids) in the insulation which when of a certain size start to producing partial discharge. At this stage not only the dielectric losses are increased but also, where high voltage are concerned, the service life of the cable may be reduced.

For distribution and transmission purposes, impregnated paper cables have had an impressive record of reliability since the turn of the century although it is being replaced these days by new types of insulation such as XLPE. The insulation in paper insulated cables consists of helically applied paper tapes. The cables designed with a belt of insulation over the laid up cores are the most common type used up to 11 kV. Beyond 11 kV, screened cables have to be used. Screening consists of a thin metallic layer in contact with the metallic sheath.

The perfect belted cable would be so manufactured that the impregnant would completely fill the interstices between the conductor, the fiber of papers and the filler material. In short, the whole volume contained within the lead or aluminum sheath would be completely void free. During installation, however, the perfect belted cable undergoes mechanical manipulation and movement of the cores relative to each other takes place. In

addition to temperature cycling, the lead sheath does not return to its original dimensions and consequently the interior of the sheath is no longer completely occupied, there being voids found within the cable. The voids contain low pressure gas extracted from the impregnant. Such voids are particularly hazardous when they occur within the highly stressed zones of the insulation.

The breakdown of paper insulating cable due to partial discharge occurs through the formation of carbonaceous on the insulation papers. This is generally known as treeing. The carbonaceous paths starts at an almost imperceptible carbon core and gradually spread outwards through the insulation.

7.3 Requirement for Stress Control

In the region of the termination of a cable dielectric screen, it can be seen that not only there is an increase in stress within the insulation in that region, but also there is a potential gradient along the interface between the dielectric and the surrounding space. The stress in the dielectric at the screen termination will be well above the design stress of the cable and premature failure can occur at this point. In addition, if the medium surrounding the termination is air then the stress in this area may be sufficient for the air to discharge even at working voltage. Thus in designing the termination of high voltage cable, it is necessary to be aware of both these problems and to include some form of stress control as following.

7.3.1 Stress cone

The traditional method of stress relief is the use of stress cone. The stress cone is a means of controlling the capacitance in the area of the screen termination, thereby reducing the

stress in the dielectric. The stress cone is continued beyond to the screen termination so as to reduce the potential gradient at the surface of the dielectric to a level where partial discharge will not occur.

7.3.2 High permittivity materials

Materials with relative permittivities significantly higher than the dielectric can provide excellent stress control at terminations. When materials of dissimilar permittivities are subjected to a potential gradient across their combined thickness, then the material with the lowest permittivity will be subjected to the highest stress. This is a physical phenomena which enables stress control to be achieved by high permittivity materials.

7.3.3 Resistive coating

Stress control can be achieved by the application of a resistive layer to the insulation surface at the screen termination. Ideally the layer will pas a small current and will therefore set up a linear voltage gradient along its length. However, the resistivity of the material has to be within quite a narrow band for the termination to work successfully. If the resistivity is too low, then the material will simply act as an extension of the dielectric screen and a high stress region will be created at the end of the resistive layer. If the resistivity is too high, then the material will have no appreciable effect and the screen termination will remain a high stress region.

7.4 Measurement of Partial Discharge in Cables

Partial discharge measurements in capacitors and short cable lengths, that behave

essentially as lumped capacitances, are straight forward. With longer cable lengths, pulse superposition due to reflected discharge pulses may lead to considerable error in the measurement of both the discharge amplitude and the discharge repetition rate. It is only with very long cables, that amplitude superposition errors due to reflection do not arise; however, the pulse counting errors remain [Bartnikas, 1990]. Partial discharge pulse reflection may be easily eliminated by impedance termination of the far cable end with resistance R in series with a partial discharge free capacitor C. Here R corresponds to the characteristics impedance of the cable and C is equal in value to the blocking capacitor used in the RLC detection circuit.

7.5 Experimental Work

To investigate the ability of texture analysis algorithms for discrimination between partial discharge sources, a 11 kV belted cable of $5m$ length was used. Sine the cable sample was very short, the cable end did not require any impedance termination and the far end was open circuit. In such sample, the following sources of partial discharge were expected:

- Internal discharge in the voids between the insulating paper
- Surface discharge at both the ends between the core and lead sheath.

The partial discharge at the cable ends could be eliminated by using stress cone. Two steel stress cones were connected to the lead sheath. Both ends of the cable were kept in the vertical position and the two cones were filled with transformer oil. Therefore, from the cable sample, the following classes of partial discharge could be measured:

- Internal plus surface discharge at a time.
- Internal discharge only, by using stress cone to suppress the surface discharge.

The high voltage source was connected to the core and the lead sheath was grounded through *RLC* measuring impedance. The measurement set up, shown in Fig. 7.1, has been used to measure the partial discharge pulses in this cable.

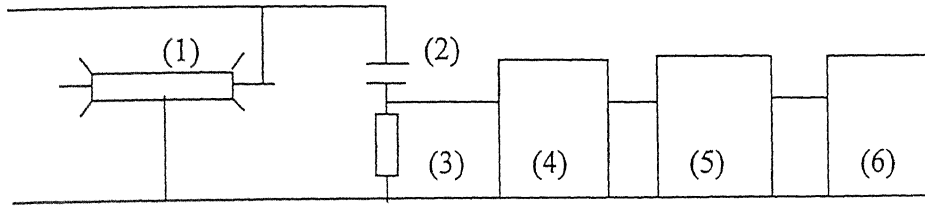


Fig. 7.1: partial discharge measurement circuit

Where,

(1) Cable test sample

(2) Coupling capacitor

(3) Measuring impedance

(4) Filter and Amplifier

(5) Digital storage oscilloscope

(6) Personal computer

7.6 Results and Discussion

Since the features of GLDHM have shown a good classification accuracy compared to the other techniques, as shown in the previous chapters, these have been used in this study for three different values of power frequency cycles which are 2, 4 and 8 cycles. The classification accuracy results are shown in Fig. 7.2. Thirty patterns were used to represent each partial discharge class. Fifteen patterns were used for training of the minimum

distance classifier and the complete set of patterns were used for testing. In case of 2 cycles, the classification accuracy of all the texture features was found to be 100%. The classification accuracy of the polarity factor (features 6 & 17, not texture feature) were 86.66% & 90% in the positive and negative half cycles, respectively and its classification accuracy increased with increasing number of cycles. However, the classification accuracy of the texture features slightly decreased for the cases of 4 and 8 cycles. The features related to the positive half cycles are found better than those related to the negative half cycle. From Fig. 7.2, one can notice the following:

(i) In the case of 2 cycles, all the texture features related to the horizontal and vertical directions have the same classification accuracy when used in the positive and negative half cycles.

(ii) In the case of 4 cycles, the classification accuracy of the contrast (feature number 14 in the horizontal direction and number 20 in the vertical direction, both in the negative half cycles) has the worst classification accuracy. However, it has the best classification accuracy in the positive half cycle (feature 3 in the horizontal direction and feature 9 in the vertical direction). In the positive half cycle, feature number 1 (mean) has almost the same classification accuracy of feature number 3 (contrast) in the horizontal and vertical directions.

(iii) Using 8 cycles, the contrast has the worst classification accuracy in the horizontal and vertical directions of both the positive and negative half cycles (features 3, 9, 14 and 20).

(iv) Also, for the case of 8 cycles, the classification accuracy of the five texture

features of GLDHM have the same relative classification accuracy when used for the horizontal or vertical direction of each half cycle individually (features 1, 2, 3, 4, 5 and 7, 8, 9, 10, 11 for the positive half cycle as well as features 12, 13, 14, 15, 16 and 18, 19, 20, 21, 22 for the negative half cycle). The same can be observed also in the case of 4 cycles in the negative half cycle only.

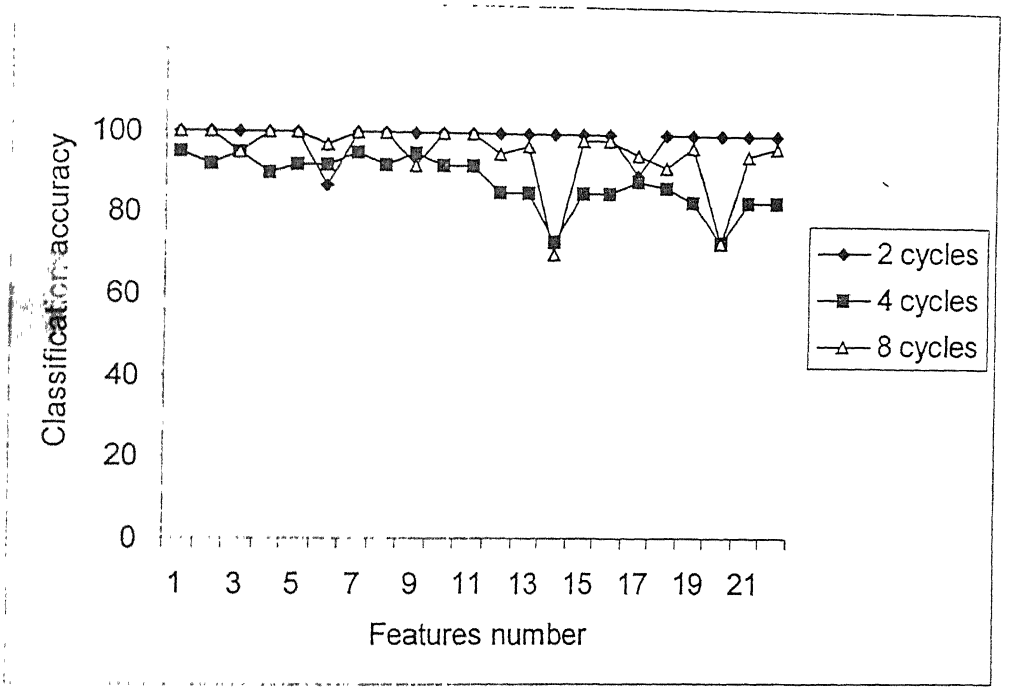


Fig. 7.2: Classification accuracy of GLDHM features for cable sample

Using two features at a time, the over all accuracy has improved for the different number of cycles. It is important to investigate the direction of both the features in each combination to determine which direction has more classification accuracy. The expected combinations will be from horizontal-horizontal features, vertical-vertical features and mixed features i.e. horizontal- vertical or vertical horizontal. Fig. 7.3 shows the directions

of the features in all the possible combinations for classification accuracy greater than certain values. These values are 100%, 95%, 90%, 85%, 80%. The figure is divided into three portions for the cases of 2, 4, and 8 cycles. The number of combinations in each direction was expressed as a percentage of the total number of possible combinations. In case of 2 cycles, the classification accuracy of all the combinations are 100% except the combination 7-17 which has 96.66% classification accuracy. In the case of 4 cycles, there is no combination having 100% classification accuracy. The maximum classification accuracy was 98.33% for about 15% from the combinations. For the case of 8 cycles, the maximum classification accuracy was 100% for around 30% from the combinations. From Fig. 7.3, the following observations can also be made:

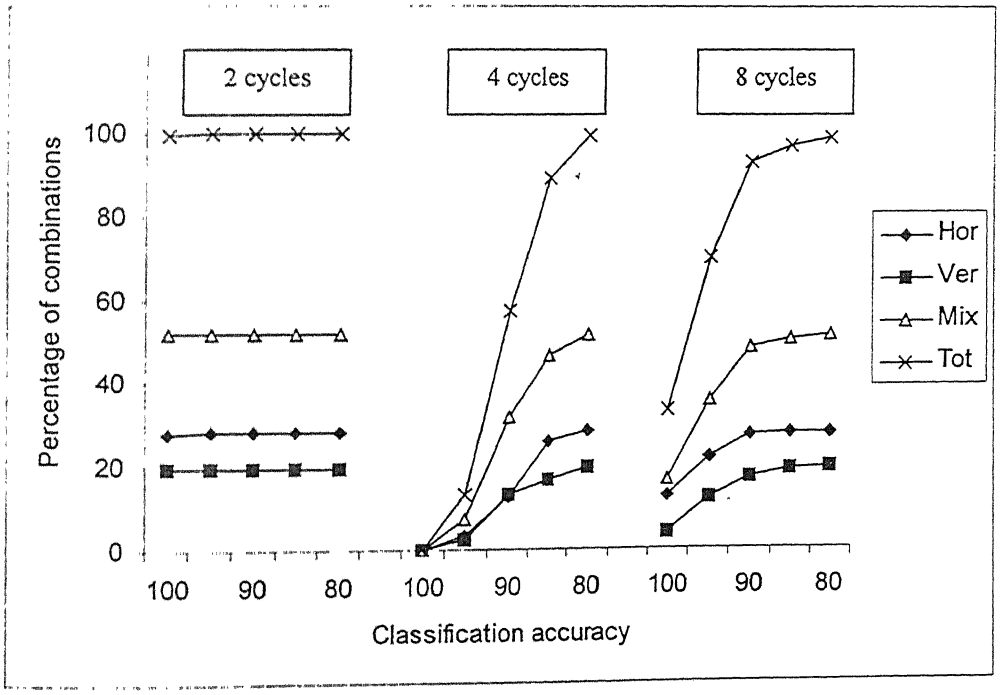


Fig. 7.3: Relation between the classification accuracy of the feature combinations of GLDHM and the direction of these features

(i) For any number of power frequency cycles, the combinations from horizontal - horizontal directions, offering a certain classification accuracy, are greater than the combinations from vertical-vertical direction.

(ii) The number of mixed combinations, offering any classification accuracy, are always greater than the summation of combinations of horizontal-horizontal and vertical-vertical directions.

(iii) In the case of 2, 4 and 8 cycles, more than 99% combinations have classification accuracy greater than 80%.

7.7 Conclusion

In this chapter, the GLDHM has been used to distinguish two different partial discharge sources generated from a 11kV belted cable sample. The first partial discharge source was a combination of the internal discharge in the voids between the insulating paper and the surface discharge at the terminal points of the cable between the core and the lead sheath. The second partial discharge was a pure internal discharge after eliminating the surface discharge by using a stress cone at both ends of the cable. The features used were able to distinguish two different partial discharge sources successfully. The features related to the horizontal direction have better discriminating power compared to the features related to the vertical direction. However, the relative classification accuracy of the texture features, when applied for the cable sample, are different than their relative classification accuracy when applied to the six partial discharge sources in chapter 4. This means that the relative classification accuracy of the texture features depends upon the partial discharge sources used to generate these features.

Chapter 8

Partial Discharge Patterns Classification Using the Principal Component Transformation

8.1 Introduction

In chapters 4 to 7, the four texture analysis methods were used to generate various features to identify different partial discharge sources. The main aim of the work carried out in these chapters was to identify the features which have the most discriminative power for the accurate classification accuracy of different partial discharge sources. From the results of the minimum distance classifier in chapter 4, it was found that the discriminative power of any feature of the four texture analysis algorithms depends upon the number of power frequency cycles used to represent the partial discharge sources. A similar result has been achieved by using the transformed divergence analysis in chapter 5. In chapter 6, the multi layer feed forward artificial neural network was used. However, it was not able to improve the classification accuracy of the texture features. In chapter 7, the features of GLDHM were used to classify two different partial discharge sources in cable sample by using minimum distance classifier. It was observed, in the previous chapters, that the relative classification accuracy of the texture features also depend upon the number of power frequency cycles. On the other hand, the relative classification accuracy of the features differed from the results obtained from the minimum distance classifier in chapter 4 for the

six partial discharge sources. In view of these observations, it is difficult to select the best feature or the best combination of features for any texture analysis by the direct features selection techniques.

The direct feature selection is done by eliminating the features which contribute little to the separability of the partial discharge sources as described in the previous chapters. Alternatively, one can use a combination of all the features by transforming the original set into a new set of features in which separability is higher in a subset of the transformed features than in any subset of the original data. The new set of features could be generated from the original set by using the principal component transformation [Jolliffe, 1986]. The principal component transformation maps the original set of features into a new and uncorrelated set of features. Moreover, it produces a space in which the data has maximum variance along its first axis, the next largest variance along a second mutually orthogonal axis and so on. The later principal components would be expected, in general, to show a smaller variance. These could be considered, therefore, to contribute little to the separability and could be ignored, thereby reducing the essential dimensionality of the classification space and thus improving classification speed. This is only of value, however, if there is correlation between the features in the original feature space and some features which have higher variance compared to the other features. The principal component transformation has already been applied to the partial discharge problem but using statistical parameters like skewness, kourtosis, etc. describing the shapes of (q_{av}-φ), (q_{max}-φ), (n-φ) distributions [Krivada, 1995].

The main aim of the work carried out in this chapter is to use the principal component transformation technique for mapping the original set of features to a reduced set in order

to minimize the classification time. The measurements of the six partial discharge sources discussed in chapter 3 and the two types of partial discharge in the cable sample described in chapter 7 have been used in the present work, for the sake of comparison. Investigate have been carried out to establish the ability of the principal component transformation to solve the problems with direct feature selection technique mentioned earlier.

8.2 Principal Component Transformation

In N dimensional space, where N is the number of features used for classification, consider that each class i.e. partial discharge source has been represented by M patterns. In this case, each pattern can be described by a vector of N components. The mean value of each component can be calculated as following:

$$mean_x(i) = \frac{1}{m}\sum_{j=1}^{m} x(i,j) \qquad \text{for i} = 1, \ldots\ldots\ldots\ldots, \text{N} \qquad (8.1)$$

While the mean values vector is useful to define the average value of the given data, it is also very important to investigate the scatter or spread of the data in the N dimensional space. The covariance matrix can be used for this purpose. Its elements $\sigma_x\ (i,k)$ are calculated as following:

$$\sigma_x(i,k) = \frac{1}{m-1}\sum_{j=1}^{m}(x(i,j) - mean_x(i))(x(k,j) - mean_x(k))^t$$

$$\text{for i} = 1,\ldots, \text{N and k} = \text{i},\ldots,\text{N} \qquad (8.2)$$

The covariance matrix provides an important mathematical concept in the analysis of data in multidimensional space. If there is a correlation between the response of partial discharge sources using certain features, the corresponding off-diagonal element in the covariance matrix will be large as compared to the diagonal elements. On the other hand, if

there is little correlation, the off-diagonal elements will be close to zero. This behavior can be described in terms of the correlation matrix R whose elements are related to the covariance matrix by,

$$R(i,j) = \sigma(i,j)/\sqrt{\sigma(i,i)\sigma(j,j)} \qquad \text{for i = 1,.....,N and k = i,...,N} \qquad (8.3)$$

It is clear that the correlation matrix is more suitable to find the correlation between the different features than the covariance matrix. Since the correlation matrix is normalized by the standard deviation of each feature, it is independent of the measuring units. It is fundamental to the development of the principal components transformation to ask whether there is a new co-ordinate system in which the data could be represented without correlation. In other words, the covariance matrix in the new co-ordinate system should be diagonal. This means that, if the vectors x describing the patterns in the original feature space are represented as vectors y in the new co-ordinate system, it is desirable to find a linear transformation G of the original co-ordinates, such that $y = G\,x$, subject to the constraint that the covariance matrix of the data in y space is diagonal. In y space the covariance matrix is, by definition,

$$\sigma_y(i,k) = \frac{1}{m-1}\sum_{j=1}^{m}(y(i,j) - mean_y(i))(y(k,j) - mean_y(k))^t \qquad (8.4)$$

and $mean_y$ is the mean vector expressed in terms of the y co-ordinate, is defined as,

$$mean_y(i) = \frac{1}{m}\sum_{j=1}^{m} y(i,j) = \frac{1}{m}\sum_{j=1}^{m} G \times x(i,j) = G \times \frac{1}{m}\sum_{j=1}^{m} x(i,j) = G \times mean_x(i,j)$$

$$\text{for } i = 1,.....,N \qquad (8.5)$$

where $mean_x$ is the data mean in x space. Therefore,

$$\sigma_y(i,k) = \frac{1}{m-1}\sum_{j=1}^{m}(G \times x(i,j) - G \times mean_x(i))(G \times x(k,j) - G \times mean_x(k))^t \qquad (8.6)$$

which can be written as

$$\sigma_y(i,k) = G \times \frac{1}{m-1} \sum_{j=1}^{m} (x(i,j) - mean_x(i))(x(k,j) - mean_x(k))^t \times G^t \qquad (8.7)$$

or

$$\sigma_y = G \times \sigma_x \times G^t \qquad (8.8)$$

Since σ_y must be diagonal, G can be recognized as the transposed matrix of eigen vectors of σ_x, provided that G is an orthogonal matrix. As a result, σ_y can then be identified as the diagonal matrix of eigen values of σ_x i.e. σ_y will be in the form of:

$$\sigma_y = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 & \ldots\ldots\ldots\ldots & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & \ldots\ldots\ldots & \ldots\ldots 0 \\ 0 & 0 & \lambda_3 & 0 & \ldots\ldots\ldots\ldots & & \ldots 0 \\ 0 & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ 0 & 0 & 0 & 0 & 0 & \ldots\ldots\ldots\ldots\ldots & \lambda_N \end{bmatrix}$$

Its elements will be the variance of the data in the respective transformed co-ordinates. It is arranged such that $\lambda_1 > \lambda_2 > \lambda_3 > \ldots\ldots > \lambda_N$. Hence the data exhibits maximum variance in y_1, the next largest variance in y_2 and so on, with minimum variance in y_N.

8.3 Eigen Value Calculation

To calculate the eigen values λ of the covariance matrix A, the following equation has to be solved

$$Ax = \lambda x \qquad (8.9)$$

Because of the presence of the unknown vector x on both sides of the above equation, the

solution methods for eigen value problem will be essentially iterative in nature. There are three different approaches to solve the eigen value problem [Griffiths, 1991]. The first method is to find the roots of the characteristic polynomial. The second one is the transformation method in which the matrix A is transformed into a new matrix B which has the same eigen values as A. However, the eigen values of the transformed matrix are easier to compute than the eigen values of the original matrix. This transformation can be done if the equation (8.9) is post multiplied by a rotation matrix and pre multiplied by the inverse of the rotation matrix to diagonalize the original matrix A. A third method is the vector iterative. In this method a guess is made for x on the left hand side of the equation (8.9), the product Ax is formed, and compared with the right hand side of the same equation. The guess is then iteratively adjusted until agreement is reached. This method is sometimes also called as power method. The major advantage of this method is that it gives the eigen value and its eigen vector directly at a time. This method always converge to the largest absolute eigen value and its corresponding eigen vector. To find the second largest eigen value of the system, the first largest value has to be removed from the system of equation once it has been computed. This step is called as deflation. Since the covariance matrix is symmetrical, its eigen vectors obey the orthogonality rules. This means that

$$x_1^T x_1 = 1 \text{ and } x_1^T x_2 = 0 \qquad (8.10)$$

where x_1, x_2 are the eigen vectors of the matrix A corresponding to the 1st and 2nd eigen values. This property could be used to established a modified matrix B such that

$$B = A - \lambda x_1 x_1^T \qquad (8.11)$$

Multiplying this equation by any eigen vector x_i to give

$$B x_i = A x_i - \lambda x_1 x_1^T x_i \qquad (8.12)$$

When i equals 1, the last equation will be zero. Thus the first eigen value of B is zero, and all the other eigen values of B are the same as those of A. Once A has deflated to B, the largest eigen value of B can be found and so on. Since the first few principal components are only required in this study, the third method has been used to determine the most important principal components.

8.4 Experimental Results

After obtaining data from the six types of PD sources in the laboratory, for each PD source, 30 patterns were generated. Out of these, 15 patterns were used for training of the minimum distance classifier while all the patterns were used for testing the classification accuracy after the training. The experimental results are described below.

8.4.1 Variance and correlation analysis

Before starting the principal component transformation, the variance and the correlation between the features were checked to determine whether the principal components exist. Figs. 8.1 to 8.4 show the variance of all the features of the four techniques GLDHM, SGLDM, GLRLM and PSM. It is clear from these figures that, with each technique, there are some features which have higher variance as compared to the other features. Such features will represent principal components as these will have higher discriminating power. It is interesting to note that these features will represent separate principal components only if there is no correlation between them. Otherwise, the principal components will be a combination of these features and each one of them will contribute to the principal components according to its relative variance. In that case, the number of principal components will reduce.

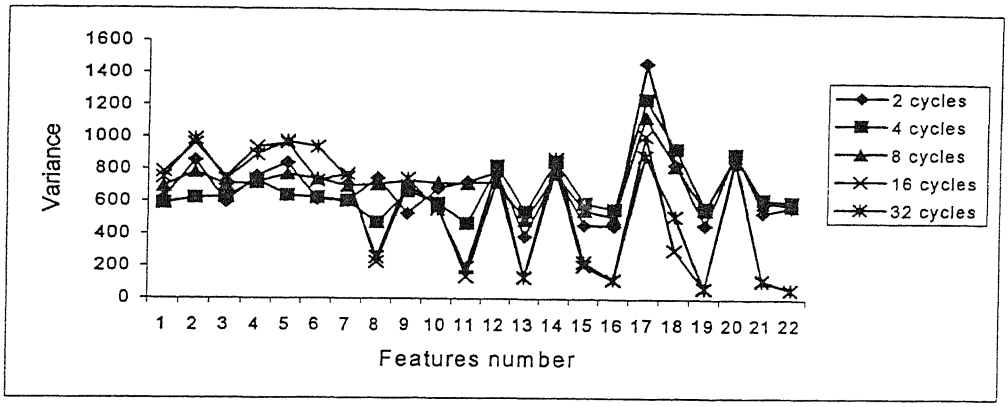


Fig. 8.1 : Variance of GLDHM features

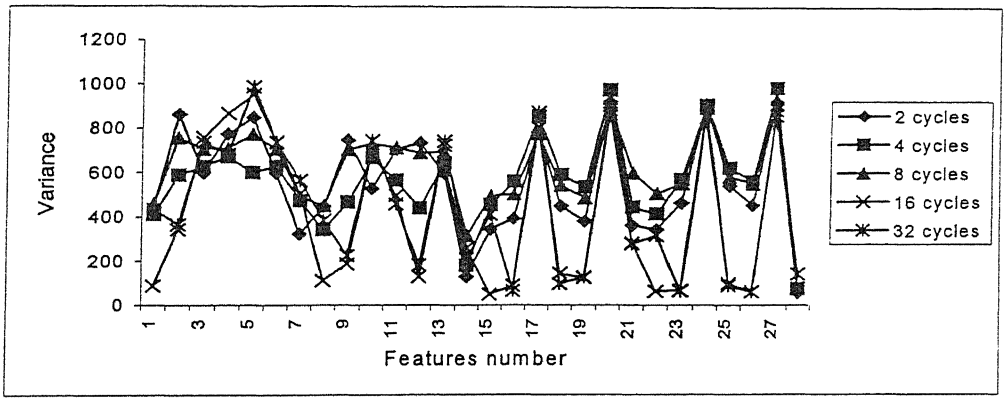


Fig. 8.2 : Variance of SGLDM features

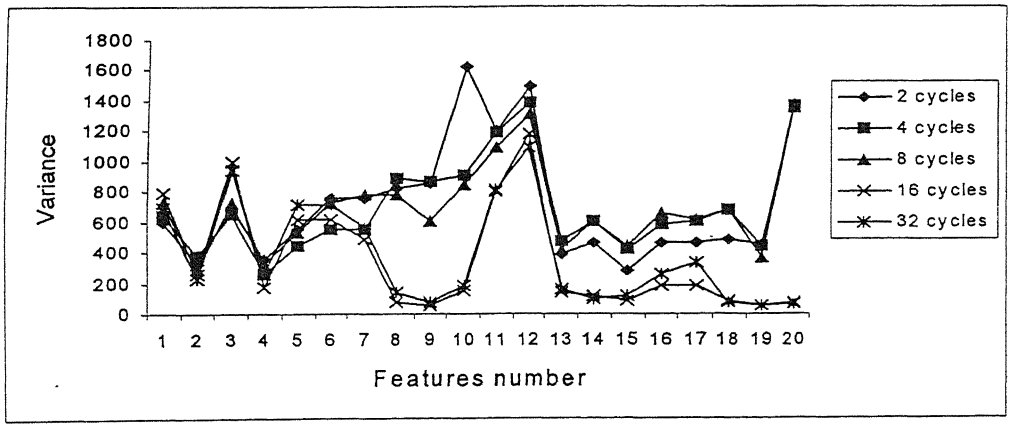


Fig. 8.3 : Variance of GLRLM features

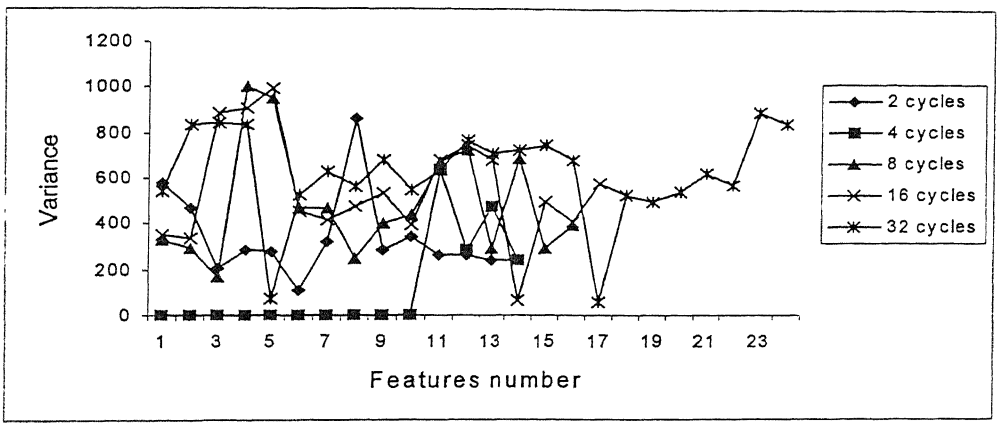


Fig. 8.4 : Variance of PSM features

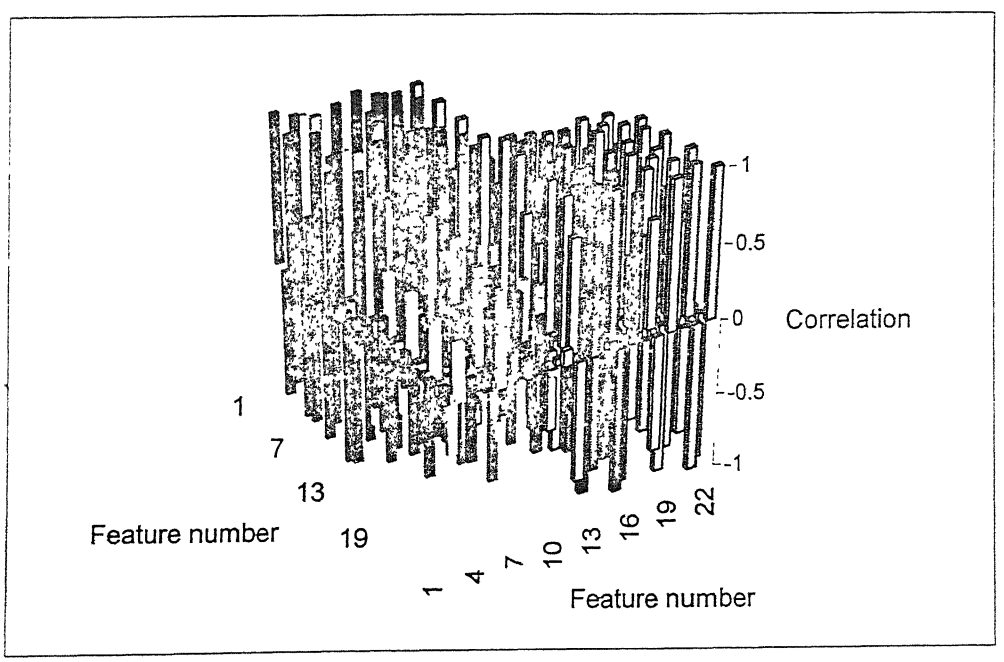


Fig. 8.5 : Correlation between the GLDHM features

Investigating the correlation between the GLDHM features, as shown in Fig. 8.5, it was found that there are high correlation between the features which describe the positive half cycles. Also there are high correlation between the features which describe the negative half cycles. On the other hand, there is no correlation between the features which describe

both positive and negative half cycles. Therefore, the expected number of principal components will be smaller as compared to the original number of features. In case of GLDHM and SGLDM, there are some features which give the maximum variance and do not depend on the number of cycles like features 12, 14, 17, 18 and 20 for GLDHM and 17, 20, 24 and 27 for SGLDM. For GLRLM such features are not found except the feature number 12. For PSM such features are not found at all.

8.4.2 Partial discharge data transformation

After verifying the possible existence of principal components in the features of each technique, the principal component transformation were carried out for the features of the four techniques individually. Since the discriminating power of PSM was poor compared to the other techniques, as shown in the previous chapters, it was felt interesting to investigate the effect of using the principal component analysis on the data generated from this technique. The distribution of the six partial discharge sources by using the first two principal components of GLDHM and PSM are shown in Figs. 8.6 and 8.7. It can be observed from these figures that the six classes are clearly separable with the two techniques, which imply that the classification accuracy by using PSM will improve with using the principal component transformation. A similar separation between the different partial discharge sources has been achieved using the other techniques. The high separability between the classes has to be reflected on the classification accuracy for the four techniques.

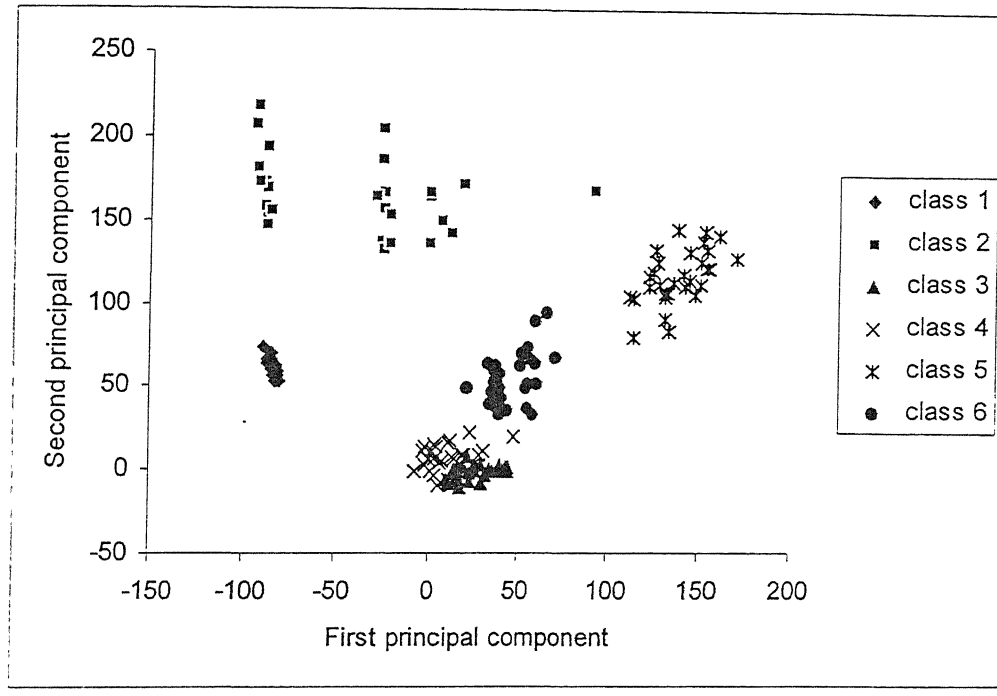


Fig. 8.6 : Distribution of partial discharge patterns using the first two principal components of GLDHM features

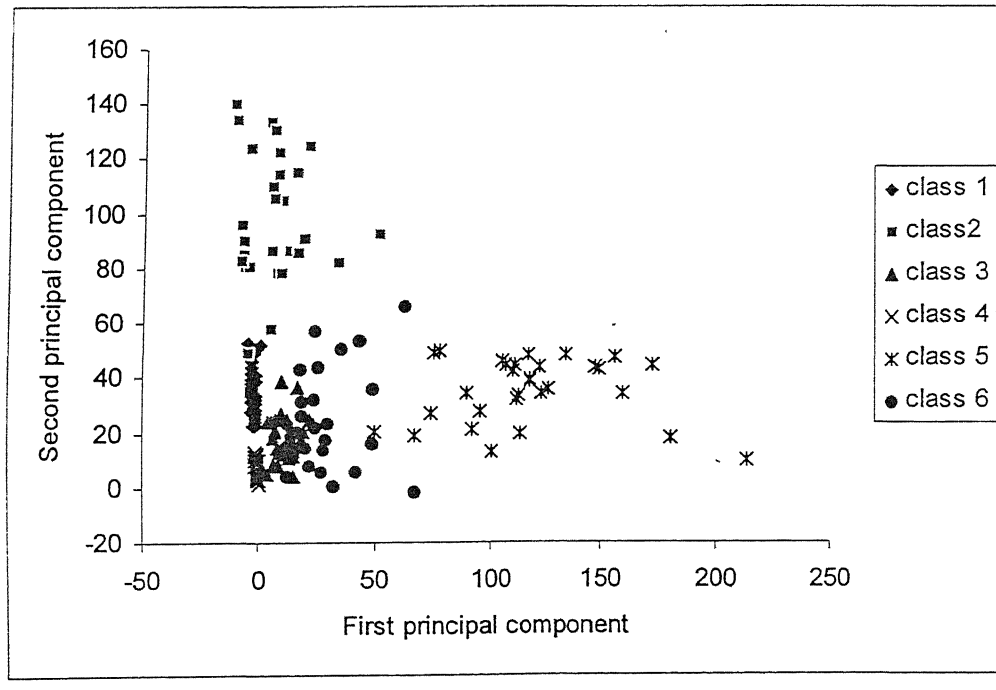


Fig. 8.7 : Distribution of partial discharge patterns using the first two principal components of PSM features

8.4.3 Classification results

Perhaps the most obvious criterion for choosing a number of principal components (say m) is to set a desired percentage of total variation which the selected principal components should contribute, say 80% or 90%[Jolliffe, 1986]. The required number of principal components is then the smallest number with which the chosen criterion is met. The definition of percentage of variation accounted for the first m principal components is

$$t_m = \frac{100 \times \sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \tag{8.13}$$

where p is the total number of the principal components.

Fig. 8.8 shows the eigen values corresponding to the first ten principal components of the features calculated from GLDHM. This figure indicates that, for any number of measured cycles, the first three principal components represent more than 90% variation in the given data. This means that only the first three principal components could be used for partial discharge classification since they have the maximum variation. This conclusion is clear from Fig. 8.9, which indicate that by using only the first principal component for the training of the minimum distance classifier, the classification accuracy achieved was around 65%. Increasing the number of principal components used for the training of the classifier to two and three components, the classification accuracy increased from 65% to more than 88% and 95%, respectively. Similar results were obtained by using the other techniques like SGLDM, GLRLM and PSM, as shown in Figs. 8.10 to 8.15. However, the classification accuracy of the principal components of the features of PSM is less than the classification accuracy of the principal components of the other techniques.

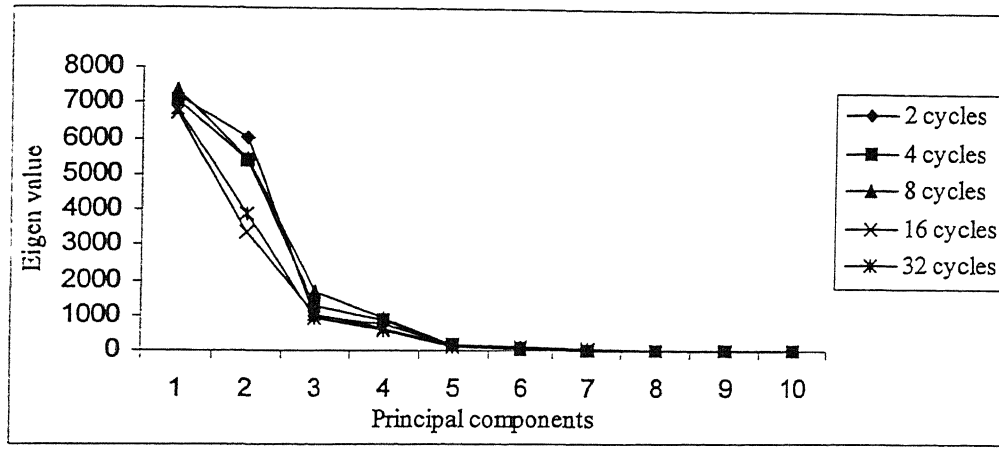


Fig. 8.8 : Eigen values of the first ten principal components of GLDHM

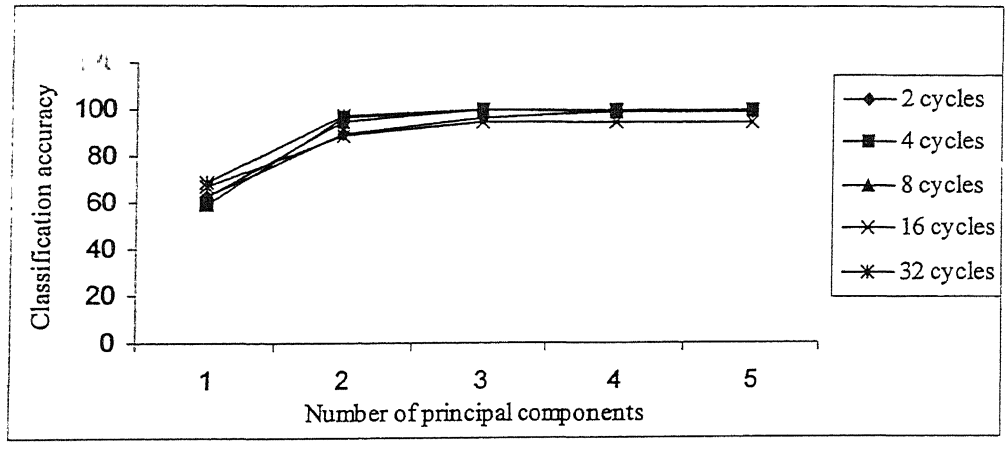


Fig. 8.9 : Classification accuracy of GLDHM principal components

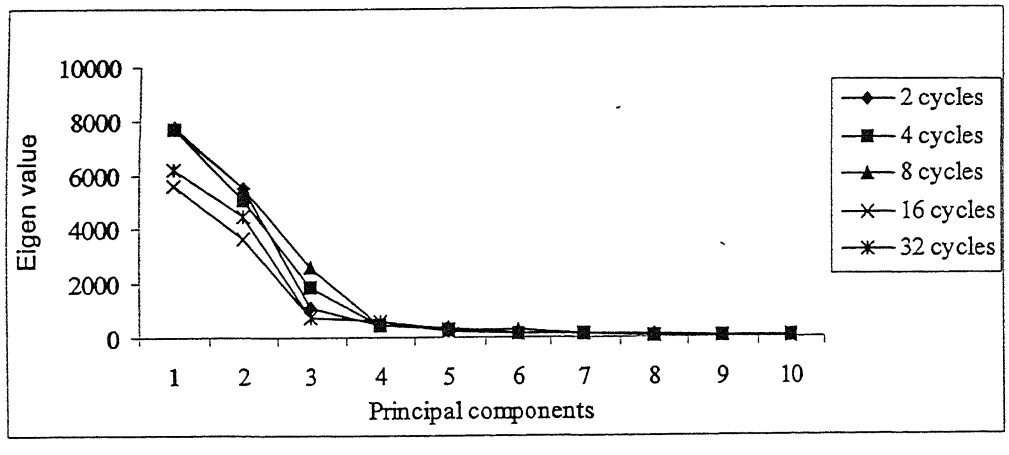


Fig. 8.10 : Eigen values of the first ten principal components of SGLDM

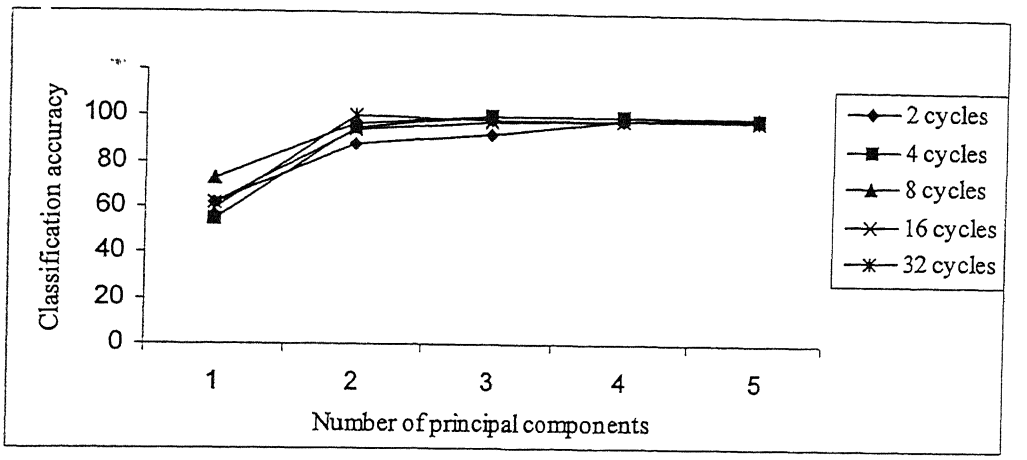


Fig. 8.11 : Classification accuracy of SGLDM principal components

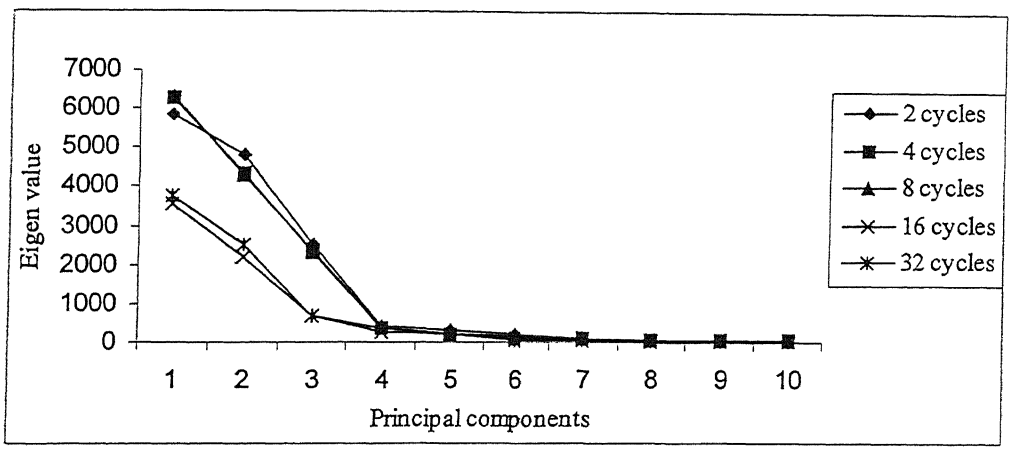


Fig. 8.12 : Eigen values of the first ten principal components of GLRLM

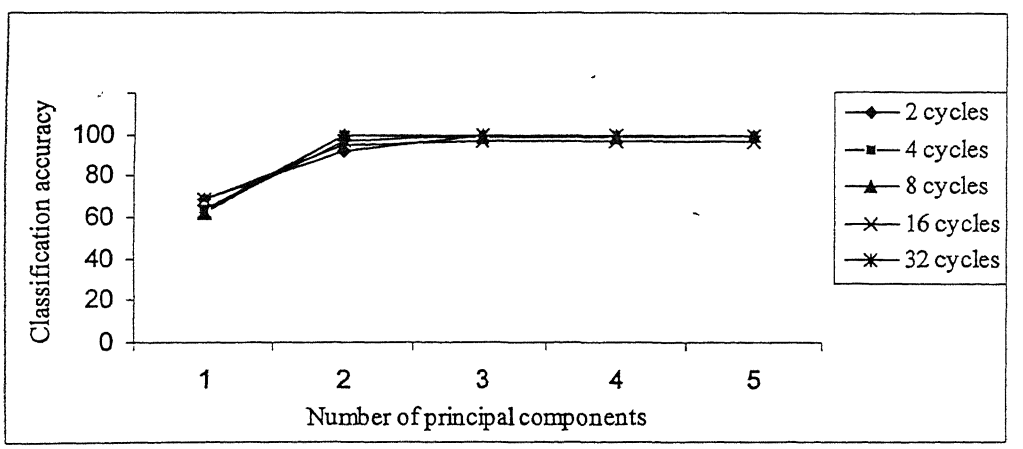


Fig. 8.13 : Classification accuracy of GLRLM principal components

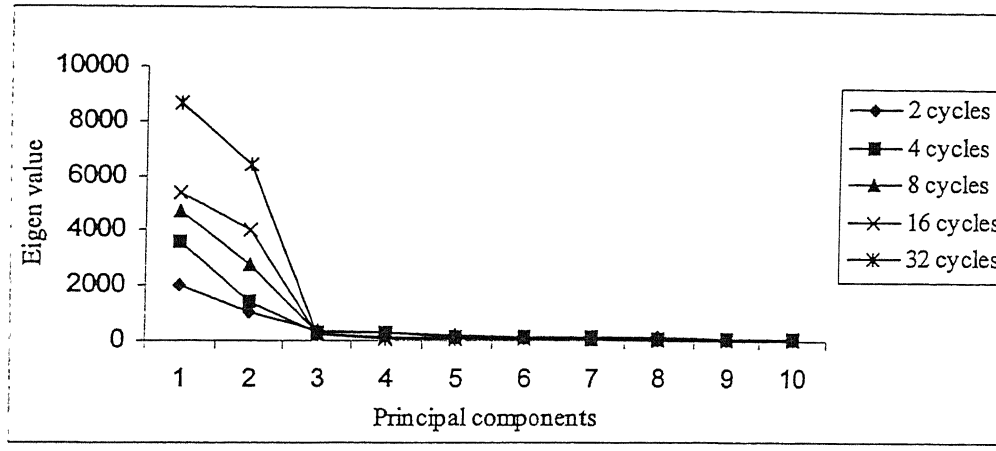


Fig. 8.14 : Eigen values of the first ten principal components of PSM

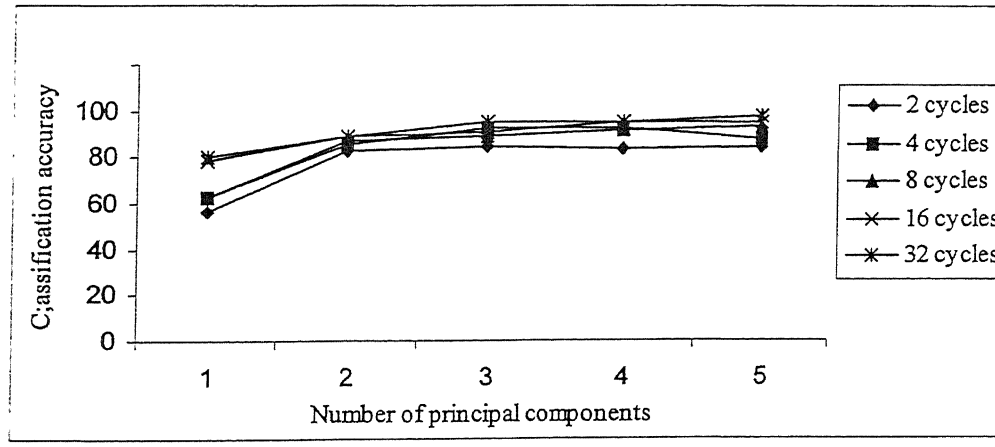


Fig. 8.15: Classification accuracy of PSM principal components

It is interesting to investigate the relationship between the first three principal components and the original set of features. The relation between the GLDHM features and their first three principal components is shown in Fig. 8.16. The summation of all squared contributions to any component was normalized to unity. By observing the correlation between the features of GLDHM which gave maximum variation, it was found that the features 12, 14, 18, 20 have almost the same variance. Also it was found that the

correlation between each pair from these features was around 97%. On the other hand, the correlation between feature 17, which has the maximum variance, and the features 12, 14, 18, 20 were less than 54%. Therefore, the contribution of the features 12, 14, 18, 20 to the first principal component were almost equal while the third principal component mainly consisted of the feature number 17 as shown in Fig. 8.16. Also it is found that the first component mainly consisted of the features 12, 14, 17, 18, 20, 21 and the second consisted of the features 1, 3, 4, 7, 9, 10 while the third consisted of the features 3, 9, 12, 14, 17, 18, 20.

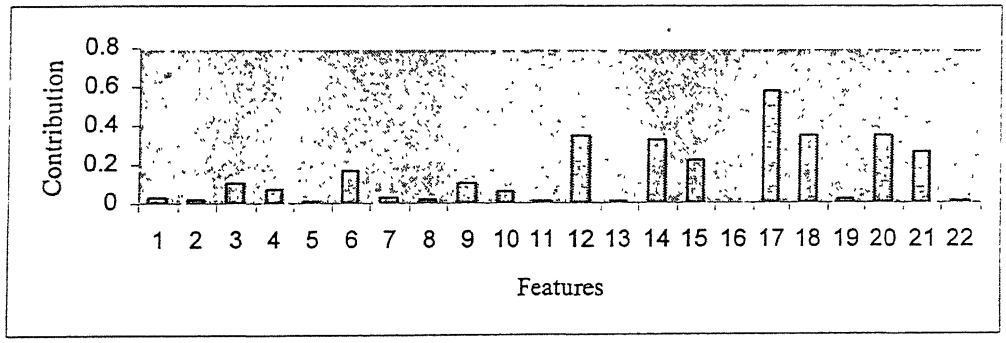


Fig. 8.16a : Contributions of GLDHM features to the first principal component

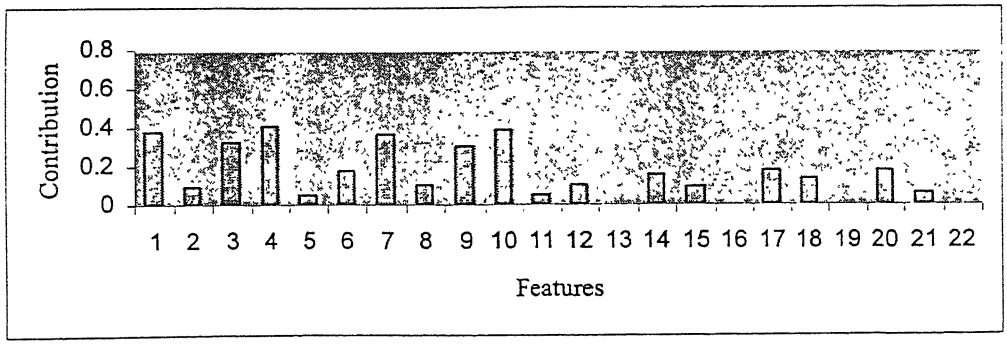


Fig. 8.16b : Contributions of GLDHM features to the second principal component

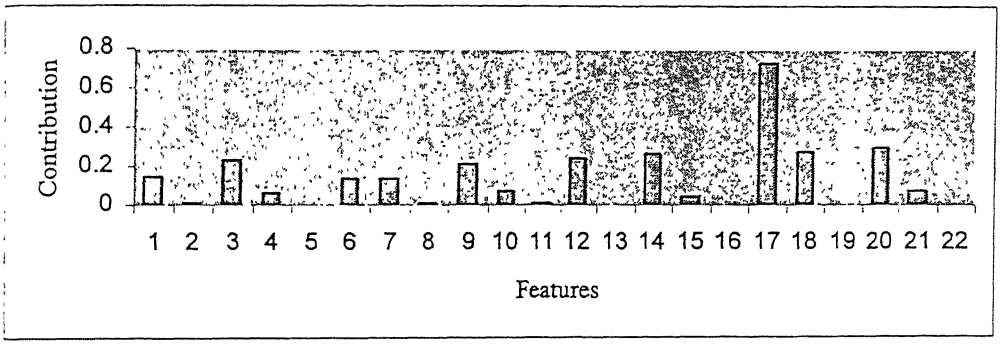


Fig. 8.16c : Contributions of GLDHM features to the third principal component

Fig. 8.16: The relation between the GLDHM features and the first thee principal components.

For SGLDM, it was found that the correlation between feature 17 and the features 20, 24, 27 were 81%, 99% and 99%, respectively, whereas the correlation between the features 20, 24, 27 were more than 98%. Fig. 8.17 shows the relation between the SGLDM features and their first three principal components. From this figure, it can be observed that the first principal component mainly consists of the features 17, 20, 24, 27 and the second principal component mainly consists of the features 3, 4, 6, 10, 11, 13, whereas the third principal component consists of the features 7, 18, 21, 25. A critical observation of Figs 8.16 & 8.17 reveals that the features contributing maximum to the first principal component, both in the case of GLDHM and SGLDM, belong to the negative half cycle whereas those contributing maximum to the second principal component belong to the positive half cycle. In both the cases, contributions of features belonging to the horizontal direction are equal to or slightly better as compared to those belonging to the vertical direction.

For GLRLM, similar results have been obtained. The first component consists of the features 11, 12, 20 and the second consists of the features 6, 7, 8, 9 whereas the third component consists of the features 10, 12, 20.

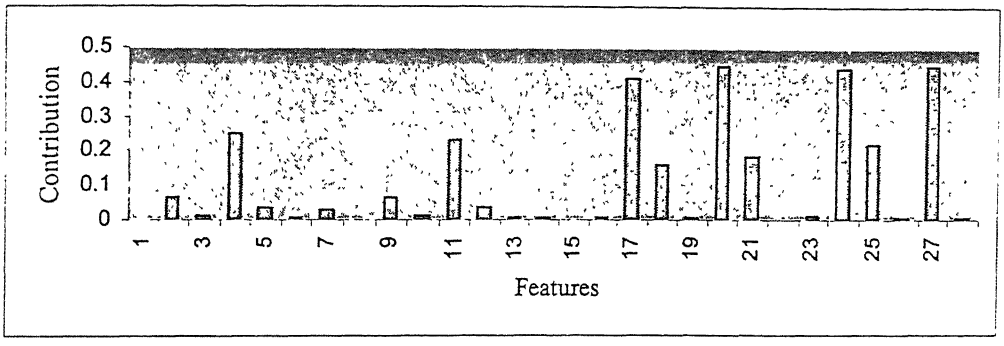


Fig. 8.17a : The contributions of SGLDM features to the first principal component

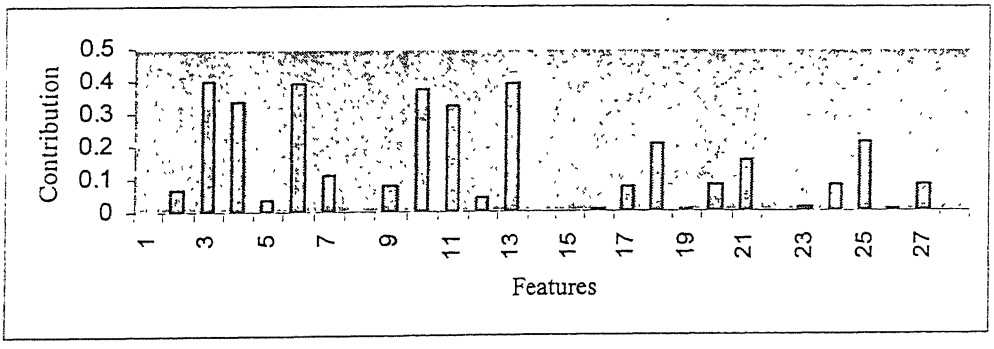


Fig. 8.17b : The contributions of SGLDM features to the second principal component

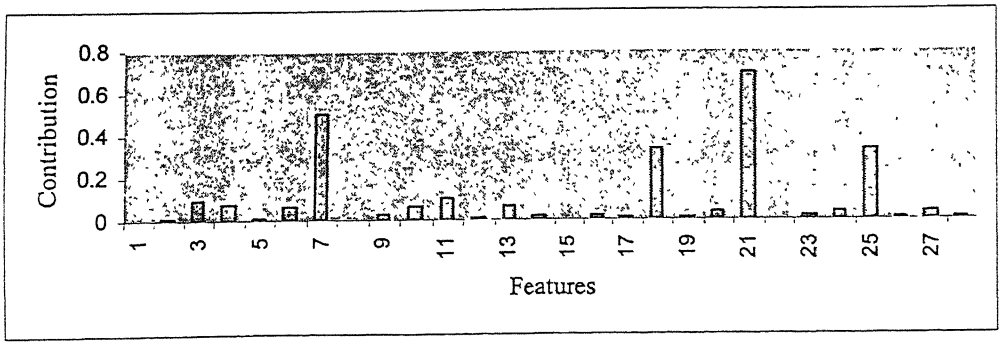


Fig. 8.17c : The contributions of SGLDM features to the third principal component

Fig. 8.17: The relation between the SGLDM features and the first thee principal components.

For PSM, the situation was different where the first component consisted of the first half of the features with almost equal contribution and nearly zero contribution from the second half of the features, while the second component was just opposite the first which mainly consisted of the second half of the features. The third one was mixed from the whole set of features.

From the above observations, it is clear that, for three of the four algorithms utilized in this work, the first principal component depends upon the features calculated from scanning the partial discharge patterns of the negative half cycle, the second is based upon the features calculated from scanning the partial discharge patterns of the positive half cycle and the third is mixed from the positive and negative half cycles. The features resulted from the horizontal direction is slightly better than the features resulted from the vertical direction.

8.5 Principal Component Transformation to the Cable Measurements

In this section, the principal component analysis has been applied to the measurements of the practical case of the cable sample described in chapter 7. The eigen values of the first six principal components of only GLDHM are shown in Fig. 8.18. From this figure, one can observe that the first principal component account for more than 95% of the total variation in the given measurements. By using the first principal component for the training of the minimum distance classifier, the classification accuracy were obtained as 100%, 93.33% and 96.66% for the cases of 2, 4 and 8 cycles, respectively as shown in Fig. 8.19. By using only two principal components the classification in the case of 4 cycles has improved to 95% while the classification accuracy for the other two cases of 2 and 8 cycles remained the same. Thus, the first two principal components are sufficient to discriminate the two sources of partial discharge in the cable.

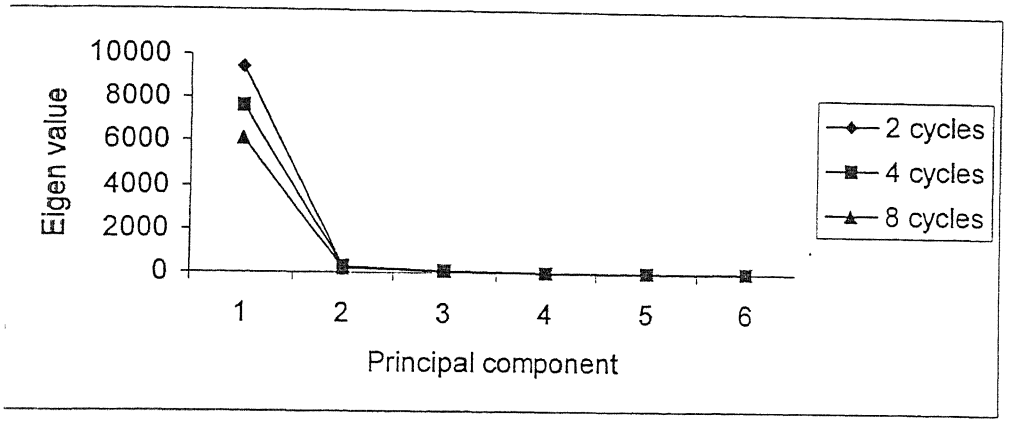


Fig. 8.18 : Eigen values of the first six principal components of GLDHM features used for the cable measurements

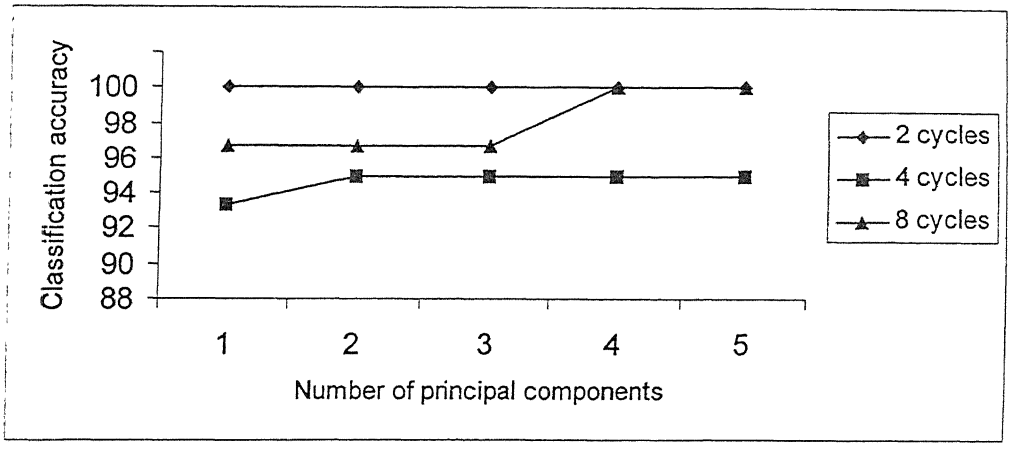


Fig. 8.19 : Classification accuracy of GLDHM principal components used for the cable measurements

Figs. 8.20a to 8.20c show the relative contribution of the various features of GLDHM belonging to the horizontal and vertical directions of the positive and negative half cycle to the first three principal components. From figures 8.20a to 8.20c, the following observations can be made:

(i) The first and the second principal components are equally divided between the positive and negative half cycles.

(ii) In each half cycle, the contribution of the features calculated from the horizontal and vertical directions are exactly the same. For example the features 1 & 7, 3 & 9 in the positive half cycle and the features 12 & 18, 14 & 20 in the negative half cycle have equal contributions.

(iii) The third principal component mainly consists of feature number 17.

The contribution of the GLDHM features to the first two principal components in this case is different than the contributions of these features for the case of the original six partial discharge sources shown in Fig. 8.16. The main reason behind this disagreement is the presence of the glow corona in the case of six partial discharge sources. Whereas, the glow corona has partial discharge pulses only in the positive half cycles, the negative half cycle will be able to classify this source very easily from the other sources. Due to this, the first principal component basically consisted of features from the negative half cycle in the case of six partial discharge sources.

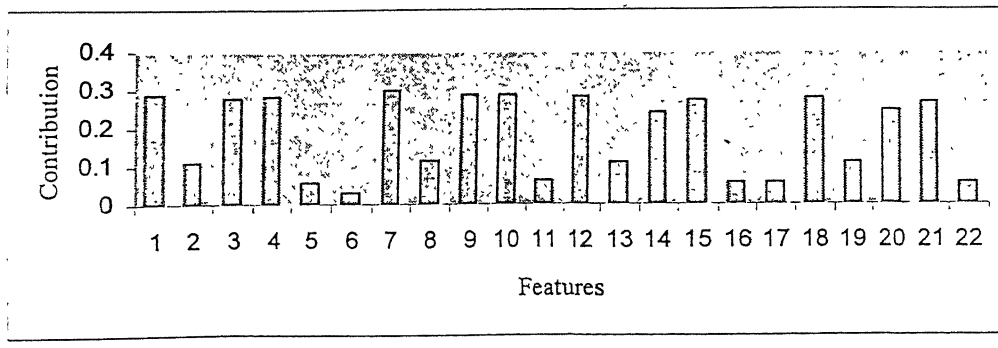


Fig. 8.20a : The contributions to the first principal component

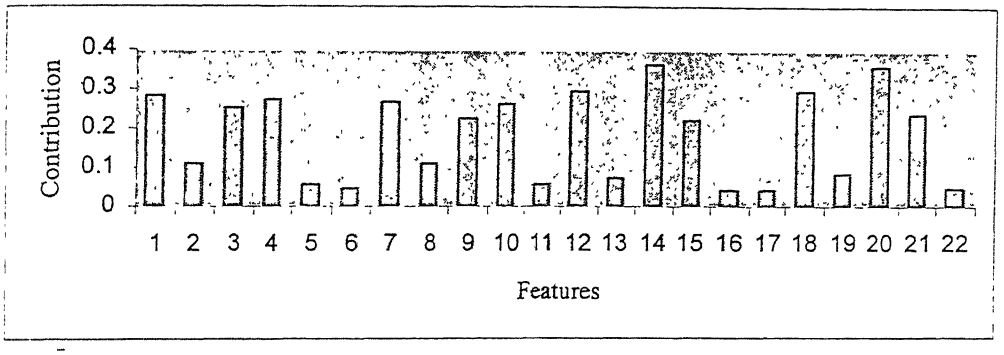


Fig. 8.20b : The contributions to the second principal component

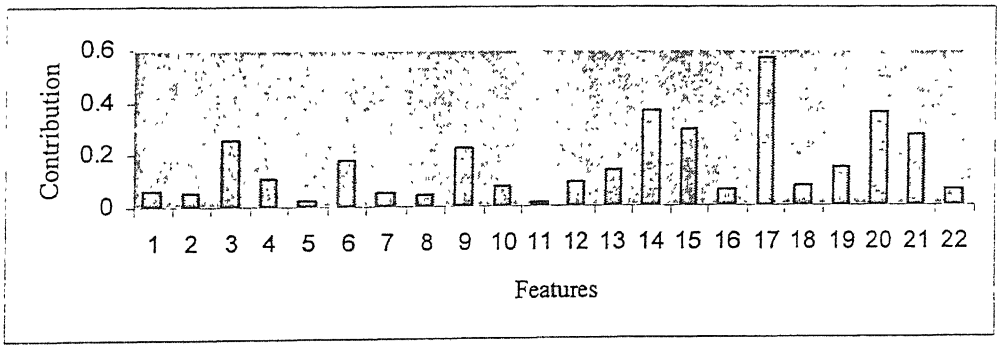


Fig. 8.20c : The contributions to the third principal component

Fig. 8.20: The relation between the GLDHM features and the first thee principal components for the cable measurements

As discussed in the preceding sections, the principal component analysis has been done for the case of the six partial discharge sources described in chapter 3 and the two partial discharge sources for the cable sample described in chapter 7. In this section, the principal component analysis for all the eight partial discharge sources has been done together. The different sources of partial discharges are now being named as class1 to class 8 which represent the glow corona, streamer corona, surface discharge, internal discharge, single protrusion, multi protrusions, internal-surface discharge in cable and the internal discharge

in cable, respectively. The largest ten eigen values are shown in Fig. 8.21. The first three principal components cover more than 90% of the over all variation of the data. When the minimum distance classifier was trained for the eight partial discharge sources, the classification accuracy were found as 54%, 90% and 97.5%, respectively, with one, two and three principal components used as shown in Fig. 8.22 for 2, 4 and 8 cycles. Hence, the first three principal components are sufficient to achieve a considerable classification accuracy. The contribution of the features to the first three principal components in the case of eight partial discharge sources are almost similar to the contribution of the features to the first three principal component in the case of the original six partial discharge sources as evident from Figs. 8.23 and 8.16. Fig. 8.24 shows the distribution of the partial discharge patterns of the eight partial discharge sources using the first two principal components of GLDHM features. It is clear from this figure that the high classification accuracy resulted from the good separation between the different partial discharge sources by using the principal component transformation.

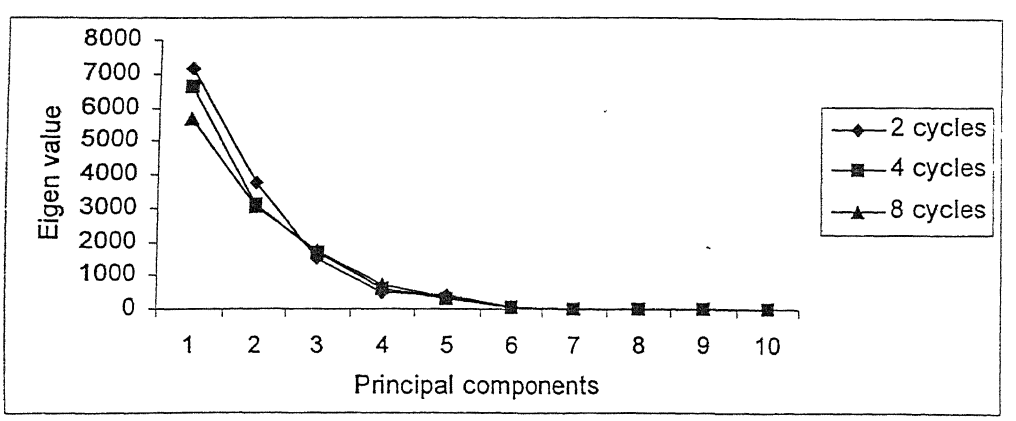


Fig. 8.21 : Eigen values of the first ten principal components of GLDHM for the eight partial discharge sources

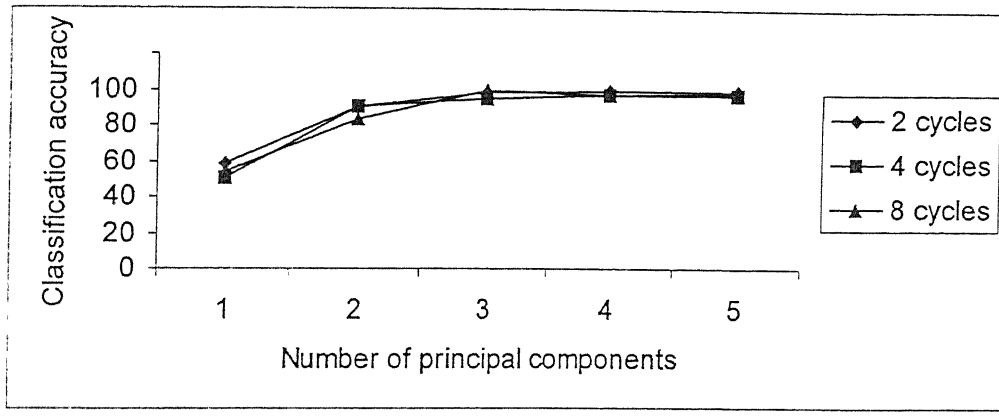


Fig. 8.22 : Classification accuracy of GLDHM principal components used for the eight partial discharge sources

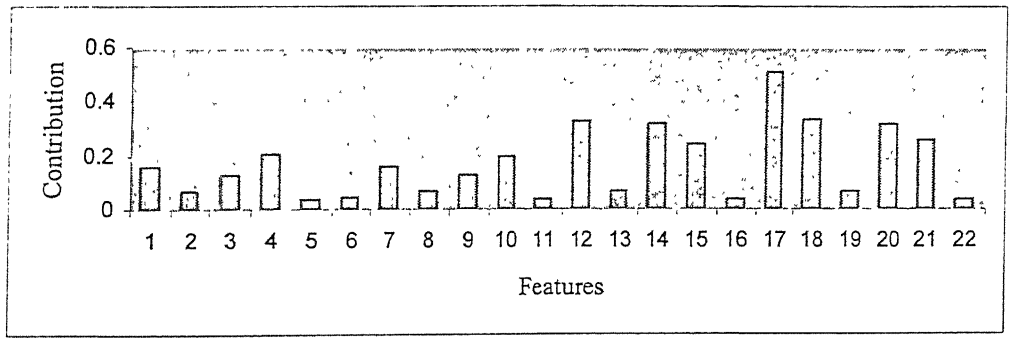


Fig. 8.23a : The contributions of GLDHM features to the first principal component for the eight partial discharge sources

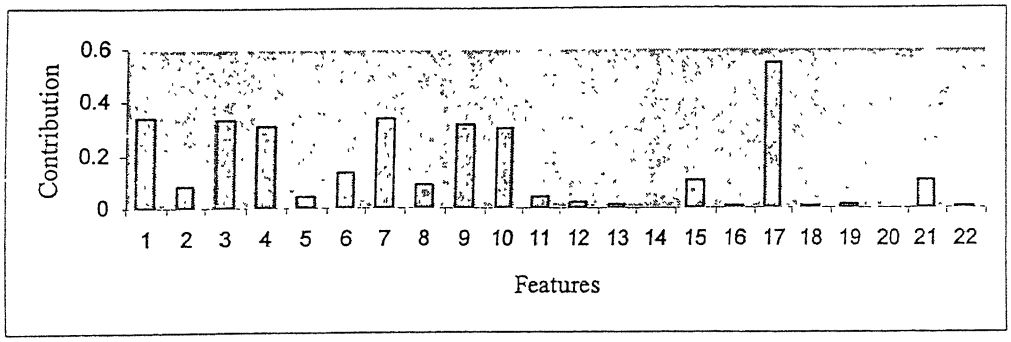


Fig. 8.23b : The contributions of GLDHM features to the second principal component for the eight partial discharge sources

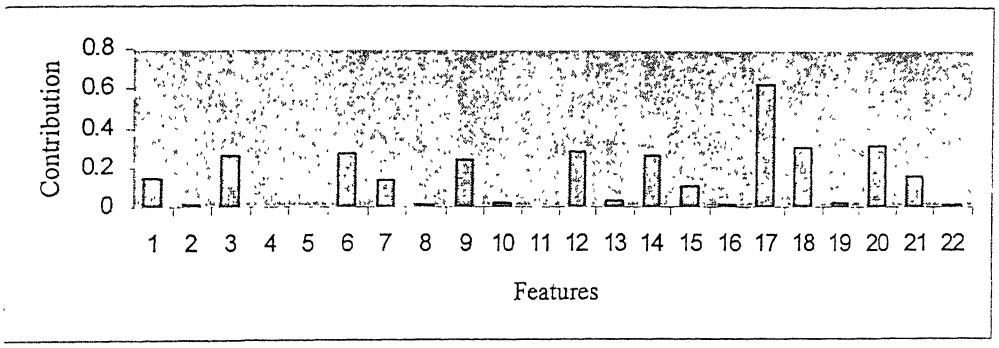


Fig. 8.23c : The contributions of GLDHM features to the third principal component for the eight partial discharge sources

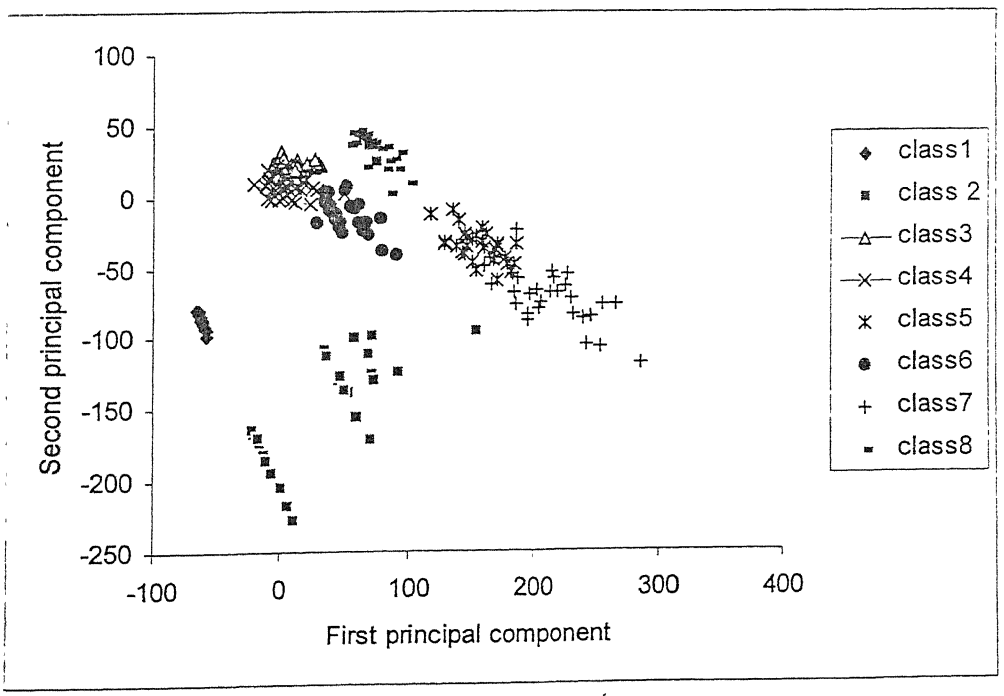


Fig. 8.24 : Distribution of partial discharge patterns of the eight partial discharge sources using the first two principal component of GLDHM features

8.6 Conclusion

In this chapter, the principal components transformation of the features based on four texture analysis techniques have been investigated to determine its ability to classify different partial discharge sources. These four texture analysis techniques are the gray level difference histogram, spatial gray level dependence method, gray level run length method

and the power spectrum method. It was found that by using the first three principal components, for any of these techniques, more than 95% classification accuracy can be achieved. Generally, the first principal component was obtained from the features which describe the negative half cycle, the second from the positive half cycles and the third from the whole set of features. The features resulting from the horizontal direction exhibit the same or slightly better performance than the features resulting from the vertical direction. The first three principal components account for more than 95% of the total variation in the data. However, the increase in the classification accuracy due to increase in the number of principal components from two to three is not significant. Therefore, in most of the cases two principal components are sufficient to achieve a considerable classification accuracy for any number of power frequency cycles used to construct the partial discharge. For each technique, changing the number of power frequency cycles used to construct the partial discharge patterns has negligible effect on the classification accuracy of the principal components. The application of principal component transformation is very useful in reducing the number of features which can be used to characterize any partial discharge source and hence to minimize the classification time. These observations are found to be valid for the six partial discharge sources created in the laboratory and also for the partial discharges in the case of the practical cable sample.

Chapter 9

Conclusions

9.1 General

Partial discharge is an electric discharge that does not completely bridge the insulation. It gives rise to electric pulses having a magnitude (q) and a phase position (ϕ) with respect to the applied voltage waveform. It has been recognized that the breakdown of insulation of an electrical equipment is often occurred by to the occurrence of partial discharge within or on the surface of the insulation. Therefore, if partial discharge is found in any insulation system, it is important to identify its source, in other words, to classify the unknown partial discharge source. The most important step in the classification process is to get the exact finger prints which can represent the different partial discharge sources successfully. During the last decade, partial discharge finger prints have been formed, commonly, by phase resolved pulse height analysis using the (q-ϕ-n) distributions, where n is the repetition rate of partial discharge pulses. Conventionally, classification of partial discharge sources has been done with the help of the shape of these distributions. However, these distributions suffer from the averaging effect and do not take the memory propagation between the partial discharge pulses into consideration.

Texture analysis algorithms have been successfully applied in the field of pattern recognition. Since the texture features contain information about the spatial distribution of

spectral variation, it is applied especially in the application of image processing. Texture analysis algorithms have been used to study the gray level variation of any image in different directions (horizontal, vertical, left diagonal and right diagonal). These algorithms could also be used to study the variation of the partial discharge pulses during the measurements. In this thesis the concept of texture analysis algorithms has been applied replacing the gray level values used in image analysis with the partial discharge pulses magnitude so that the texture features could give an over view of every single partial discharge pulse of the whole measurements. Investigations in this thesis were conducted in the horizontal and vertical directions only to study the relation between each adjacent pulses in the same cycle as well as the relation between a pulse and the other pulses in the same phase angle in different cycles. An attempt was made to understand some of the issues in obtaining reliable features from texture analysis algorithms and evaluate its ability for classifying different partial discharge sources using automated personal computer system. Four texture analysis algorithms have been studied in this work. Three of them include spatial gray level dependence method (SGLDM), gray level difference histogram method (GLDHM) and gray level run length method (GLRLM) in the spatial domain. Whereas the fourth algorithm was in frequency domain namely the power spectrum method (PSM). The work carried out in the present thesis broadly cover the following.

- The classification accuracy of the different texture analysis techniques have been calculated by using the minimum distance classifier and the discriminating power of the texture features have been compared with the conventional (q-φ-n) method .
- The relative classification accuracy of every feature in each texture analysis algorithms

have been calculated by using the minimum distance classifier to determine the maximum classification accuracy for identifying the best feature.

- Dissimilarity between different partial discharge sources was measured using the transformed divergence analysis as a direct feature selection technique to reduce the classification time by eliminating the features which contribute little to the separation between the partial discharge sources.
- The discriminating power of different features were also calculated by using an artificial neural network (ANN) as a non parametric classifier to overcome the problem of partial discharge sources which did not have normal distribution in their feature space.
- The principal component analysis was also used, as an indirect feature reduction method, to determine the number of principal components which could be used at a time to achieve reasonable classification accuracy.

The above methods were applied for classification of six partial discharge sources created in the laboratory and also for the partial discharge measured from a practical cable sample.

9.2 Main Conclusions

From the research work carried out in this thesis the following main conclusions can be drawn:

1. Texture analysis algorithms can be used to generate different features which are capable of distinguishing between different partial discharge sources.
2. The classification accuracy of features generated from SGLDM and GLDHM algorithms are much better than the features generated from the phase resolved pulse

height distributions especially for classifying unknown partial discharge patterns not used for training of the classifier.

3. PSM is computationally intensive and it uses the complex number rather than the real number. Also it needs more computer memory to save the original partial discharge as well as the transformed partial discharge patterns.
4. Most of the features of the GLDHM and SGLDM have the same discriminating power.
5. GLDHM has a classification accuracy at par with the other algorithms and is computationally faster. Therefore, it is recommended for partial discharge source classification.
6. Using the minimum distance classifier, for each of the algorithms, two features at a time are sufficient to achieve a considerable classification accuracy.
7. For each of the algorithms, employing average transformed divergence, two features at a time are also sufficient to achieve a complete separation between the different partial discharge sources.
8. However, the discriminating power of the features using the minimum distance classifier and the separation between the classes using the average transformed divergence depend on the number of cycles to construct the patterns which further generate the features. It also depends upon the partial discharge sources considered for classification.
9. The artificial neural network (ANN) is relatively time consuming process during the learning phase. Using a single feature for training, the classification accuracy of the best features were less than the classification accuracy of the best features resulted from the minimum distance classifier for all the four techniques in this work. A similar

result was achieved when two features were used at a time for the network training. The over lapping between the different partial discharge classes affects the ability of the ANN for classification.

10. For the principal component transformation, two principal components were found to be sufficient to achieve a considerable classification accuracy in most of the cases.
11. In each half cycle, the contribution of the features related to the horizontal and vertical directions are almost the same for the first and second principal components.
12. The classification accuracy of the principal components are independent of the number of cycles used to construct the partial discharge patterns.
13. The application of the principal component transformation reduces the number of features needed for partial discharge classification and hence the classification time. These observations were found equally valid for the six partial discharge sources created in the laboratory and the two partial discharge in the practical cable sample. Therefore, the principal component transformation is recommended as feature selection technique for the partial discharge source classification.

9.3 Scope for Further Work

In this thesis, the features calculated from the texture analysis algorithms have been used successfully for the classification of partial discharge sources. These features exhibit a good performance compared to the conventional pulse height phase resolved method. The features calculated from the conventional method have been reported to investigate the effect of aging on the shape of the q-φ-n distributions. It has been also reported that the shape of these distributions changes with aging of the insulation. However, it is felt that

extent of the insulation damage by observing these features can be studied in detail for monitoring the condition of a particular insulation. The features of texture analysis algorithms have proven to be superior compared to the conventional q-φ-n distribution. By observing the texture features, the assessment of the electrical insulation condition can be achieved.

In the present thesis, only multilayer perceptron model of ANN has been tried out. Some of the recent ANN models and a combination of fuzzy logic and ANN can also be tried out for the partial discharge classification.

References

Andera Baraldi and Flavio Parmiggiani (1995) "An Investigation of the Textural Characteristics Associated With Gray Level Co-occurrence Matrix Statistical Parameters" IEEE Trans. on Geoscience and Remote Sensing, Vol. 33, No. 2, pp. 293-304.

Austin J and James R. E. (1976) "On-line Digital Computer System for Measurement of Partial Discharge in Insulation Structures" IEEE Trans. on Electrical Insulation, Vol. 11, No. 4, pp. 129-139.

Bartnikas R. and John H. E. Levi (1969) "A Simple Pulse-Height Analyser for Partial Discharge Rate Measurements" IEEE Trans. on Instrumentation and Measurement, Vol. IM-18, No.4, pp. 341-345.

Bartnikas R. (1973a) "Use of Multichannel Analyser for Corona Pulse Height Distribution Measurements on Cable and Other Electrical Apparatus" IEEE Trans. on Instrumentation and Measurement, Vol. IM-22, No.4, pp. 403-407.

Bartnikas R. (1973b) "Note on Multichannel Corona Pulse height Analyser" IEEE Trans. on Electrical Insulation, Vol. EI-8, No. 1, pp. 2-5.

Bartnikas R. (1987) "A Commentary on Partial Discharge Measurements and Detection" IEEE Trans. on Electrical Insulation, Vol.22, No.5, pp. 629-653.

Bartnikas R. (1990) "Detection of Partial Discharges (Corona) in Electrical Apparatus" IEEE Trans. on Electrical Insulation, Vol. 25, No. 1. pp. 111-124.

Battiti, R. (1989) "Accelerated Back Propagation Learning: Two Optimization Methods" Complex Systems, Vol. 3, pp. 331-342.

Bhattacharyya G. K. (1977), Statistical Method Concepts and Methods, John Wiley & Sons Inc.

Bungay E. W. G. and McAllister D. (1990), Electric Cables Handbook, BSP Professional books.

Buttner G., Hajos T. and Korandi M. (1989) "Improvements to the Effectiveness of Supervised Training Procedure" International Journal of Remote Sensing, Vol. 10, No. 6, pp. 1005-1013.

Capponi G. and Schifani R. (1992) "Measurement of Partial Discharge in Solid Dielectrics with a Microprocessor-based System" IEEE Trans. on Electrical Insulation, Vol. 27, No. 1, pp.106-113.

Conners R. W. and Harlow C. A. (1980) "A Theoretical Comparison of Texture Algorithms" IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 2, No. 3 pp. 204-222.

Devore J. L. (1987), Probability and Statistics for Engineering and the Science, 2nd edition. Wadsworth, Inc.

Ehrich R. W. and Foith, J. P. (1978) "A View of Texture Topology and Texture Description" Computer Graphics and Image Processing, 8, pp. 174-202.

El-Makkawy S. M. (1998) "Insulation Behavior of SF6 Gas in Non Uniform Field Under Li and VFT Using A Neural Network" IEEE International Symposium on Electrical Insulation, Arlington, Virginia, USA, pp. 701-705

Galloway M. M. (1975) "Texture Analysis Using Gray Level Run Length" Computer Vision, Graphics, and Image Processing, 4, pp.172-179.

Gool, L. V. Dewaele P. and Oosterlinck A. (1985) "Texture Analysis Anno 1983" Computer Vision, Graphics and Image Processing, 29, pp. 336-357.

Griffiths D. V. and Smith I. M. (1991), Numerical Methods for engineers, Blackwell Scientific Publications.

Gulski E., Morshuis P. H. F. and Kreuger F. H. (1990) "Atomised Recognition of Partial Discharges in Cavities" Japanese Journal of Applied Physics. Vol.29, No.7, pp. 1329-1335.

Gulski E. and Kreuger F. H. (1992) "Computer-aided Recognition of Discharge Sources" IEEE Trans. on Electrical Insulation, Vol. 27, No. 1, pp. 82-92.

Gulski E. (1993) "Computer-aided Measurement of Partial Discharge in HV Equipment" IEEE Trans. on Electrical Insulation, Vol. 28, No. 6, pp. 969-983

Gulski E. and Krivda A. (1993) "Neural Network as a Tool for Recognition of Partial Discharge" IEEE Trans. on Electrical Insulation, Vol. 28, No. 6, pp. 984-1001.

Gulski E. (1995a) "Diagnosis of HV Components by Digital PD Analyzer" IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 2, No. 4, pp. 630-640.

Gulski E. and Krivda A. (1995) "Influence of Aging on Classification of PD in HV Components" IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 2, No. 4, pp. 676-684.

Gulski E. (1995b) "Digital Analysis of Partial Discharges" IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 2, No. 5, pp. 822-837.

Haralick, R. M., Shanmugam K. S., and Dinstein I. (1973) "Texture Features for Image Classification" IEEE Trans. on System, Man, and Cybernetics, SMC-3, No.6, pp. 610 - 621.

Haralick R. M. (1979) "Statistical and Structural Approaches to Texture" Proceeding of IEEE, Vol. 67, No. 5, pp. 786-804.

Hawkins J. K. (1970) "Texture Properties for Pattern Recognition" Lipkin B. S. and Rosenfeld A. (eds.), Picture Processing and Psychopictorics. Academic Press, New York, pp. 347-370.

Hozumi N. , Okamoto T. and Imajo T. (1992) "Discrimination of Partial Discharge Pattern Using Neural Network" IEEE Trans. on Electrical Insulation, Vol. 27, pp. 550-556.

Irons J. R. and Petersen G. W. (1981) "Texture Transforms of Remote Sensing Data" Remote Sensing of Environment, 11, pp. 359-370.

James R. E. and Phung B. T. (1995) "Development of Computer-Based Measurements and Their Application to PD Pattern Analysis" IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 2, No. 5, pp. 838-856.

Jan J. Tuma (1976), Handbook of Physical Calculations, McGraw Hill, Inc.

Jolliffe L. T. (1986), Principal Component Analysis, Springer-Verlag New York Inc.

Kreuger F. H. , Morshuis P. H. F. and Sonneveld W. A. (1988) "Optical Detection of Surface Discharge" IEEE Trans. on Electrical Insulation, Vol.23, No. 3 pp. 447-449.

Kreuger F. H. (1989) ,Partial Discharge Detection in High Voltage Equipment, Butterworths.

Kreuger F. H., Gulski E. and Krivda A. (1993) "Classification of Partial Discharges" IEEE Trans. on Electrical Insulation, Vol. 28, No. 6, pp. 917-931.

Krivda A., Gulski E. (1994) "Influence of Aging on Classification of Partial Discharges in Cavities" Japanese Journal of Applied Physics. Vol.33, No.10, pp. 5942-5949.

Krivda A. (1995a) "Automated Recognition of Partial Discharge" IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 2, No. 5, pp. 796-821.

Krivda A., Gulski E., Satish L. and Zaengl W. S. (1995b) "The Use of Fractal Features for Recognition of 3-D Discharge Pattern" IEEE Trans. on Dielectrics and Electrical Insulation Vol. 2, No. 5, pp. 352-359.

Kuffel E. and Zaengl W. S. (1984) ,High Voltage Engineering, Pergamon Press.

Lippmann R. P. (1987) "An Introduction to Computing with Neural Networks" IEEE ASSP Magazine, Vol.4, pp. 4-22.

Mausel P. W., Kramber W. J. and Lee J. K. (1990) "Optimum Band Selection for Supervised Classification of Multispectral Data ", Photogrametric Engineering and Remote Sensing, Vol. 56, No.1, pp. 55-60.

Mazroua A. A., Salama M. M. A. and Bartnikas R. (1993) "Partial Discharge Pattern Recognition with Neural Networks using Multilayer Perceptron Techniques" IEEE Trans. on Electrical Insulation, Vol. 28, pp.1082-1089.

Mazroua A. A., Salama M. M. A. and Bartnikas R. (1994) "Discrimination Between PD Pulse Shapes Using Different Neural Network Paradigms" IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 1, No. 6, pp. 1119-1131.

Mazroua A. A., Salama M. M. A. and Bartnikas R. (1995) "Neural Network System Using the Multi-layer Perceptron Technique for the recognition of PD Pulse Shapes Due to Cavities and Electric Trees" IEEE Trans. on Power Delivery, Vol. 10, pp. 92-96.

Meijer S., Gulski E. and Smit J. J. (1998) "Pattern Analysis of Partial Discharges in SF_6 GIS" IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 5, No. 6, pp. 830-842.

Okamoto Tatsuki and Tanaka Toshikatsu (1982) "Change of Pulse-Mean φ-q Distribution Patterns of Partial Discharges Due to Treeing Propagation" Electrical Engineering in Japan, Vol. 102, No. 6, pp. 18-25.

Okamoto Tatsuki and Tanaka Toshikatsu (1985) "Prediction of Treeing Breakdown from Pulse Height of Partial Discharge on Voltage-Phase Angle" Japanese Journal of Applied. Physics Vol.24, No.2, pp. 156-160.

Okamoto Tatsuki and Tanaka Toshikatsu (1995) "Partial discharge Pattern Recognition for Three Kinds of Model Electrodes with A Neural Network" IEE Proceeding of Science, Measurement and Technology, Vol. 142, No. 1, pp. 75-84.

Peter J. Tavner and James Penman (1987), Condition Monitoring of Electrical Machines, Research Studies Press ltd.

Pfaffenberger R. C. and Patterson J. H. (1977), Statistical Methods for Business and Economics, Richard D. Irwin, Inc.

Pratt W. K. (1978), Digital image processing, John Wiley & Sons, Inc.

Ramacant A. Gayakwad (1993), Op-Amps and Linear Integrated Circuits, Prentice Hall of India, New Delhi.

Richards John A. (1993), Remote Sensing Digital Image Analysis, Springer–Verlag,

Berlin.

Ronald H. Harrold (1975) "Ultrasonic Spectrum Signatures of Under-Oil Corona Source" IEEE Trans. on Electrical Insulation Vol. EI-10, No. 4, pp. 109-112.

Satish L. and Zaengl W. S. (1994) "Artificial Neural Networks for Recognition of 3-d Partial Discharge Patterns" IEEE Trans. on Dielectric and Electrical Insulation, Vol. 1, No. 2, pp. 265-275.

Satish L. and Zaengl W. S. (1995) "Can Fractal Features be Used for Recognizing 3-d Partial Discharge Patterns?" IEEE Trans. on Dielectrics and Electrical Insulation Vol. 2, No. 3, pp. 352-359.

Simon Haykin (1994), Neural Networks, Macmillan College Publishing Company, Inc.

Steiner J. P. (1991) "Partial Discharge – Part IV : Commercial PD Testing" IEEE Electrical Insulation Magazine. Vol. 7, No. 1, pp. 20-33.

Suzuki H. and Endoh T. (1992) "Pattern Recognition of Partial Discharge in XLPE Cables Using a Neural Network" IEEE Trans. on Electrical Insulation, Vol. 27, pp. 543-549.

Swain P. H. and Davis S. M. (1978) "Remote Sensing: The Quantitative Approach" McGraw-Hill, NewYork.

Tamura H., Mori S. and Yamawaki T.(1978) "Texture Features Corresponding to Visual Perception" IEEE Trans. on Systems, Man, and Cybernetics, SMC-8, No. 6, pp. 460-473.

Tou J. T. and Gonzalez R. C. (1974), Pattern Recognition Principles, Addison-Wesley.

Thomas, J. O. (1977) "Texture Analysis in Image Processing" Proceeding of British Pattern Recognition Association and remote sensing society, April, Oxford, pp. 1- 43.

Van Brunt R. J. and Kulkarni S. V. (1989) "Method for Measuring the Stochastic Properties of Corona and Partial Discharge Pulses" Review of Scientific Instrument, Vol. 60, No. 9, pp. 3012-3023.

Van Brut R. J. and Cernyar E. W. (1991) "Influence of Memory Propagation on Phase Resolved Stochastic Behaviour of ac-Generated Partial Discharge" Applied Physics Letter, Vol. 58, No. 23, pp. 2628- 2630.

Van Brut R. J., Cernyar E. W., and Von Glahn P. (1993) "Importance of Unravelling Memory Propagation Effects in Interpreting Data on Partial Discharge Statistics" IEEE trans. on Electrical Insulation, Vol. 28, No. 6, pp. 905-916.

Van Ooyen A. and Nienhuis B. (1992) "Improving the Convergence of the Back

Propagation Algorithm" Neural Networks, Vol. 5, pp. 465-471

Weszka J. S. , Dyer C. R. and Rosenfeld A. (1976) "A Comparative Study of Texture Measures for Terrain Classification" IEEE trans. on System, Man, and Cybernetics, SMC-6, No.6, pp. 269 - 285.

Appendix 1: Histogram of the Features of the Texture Analysis Algorithms

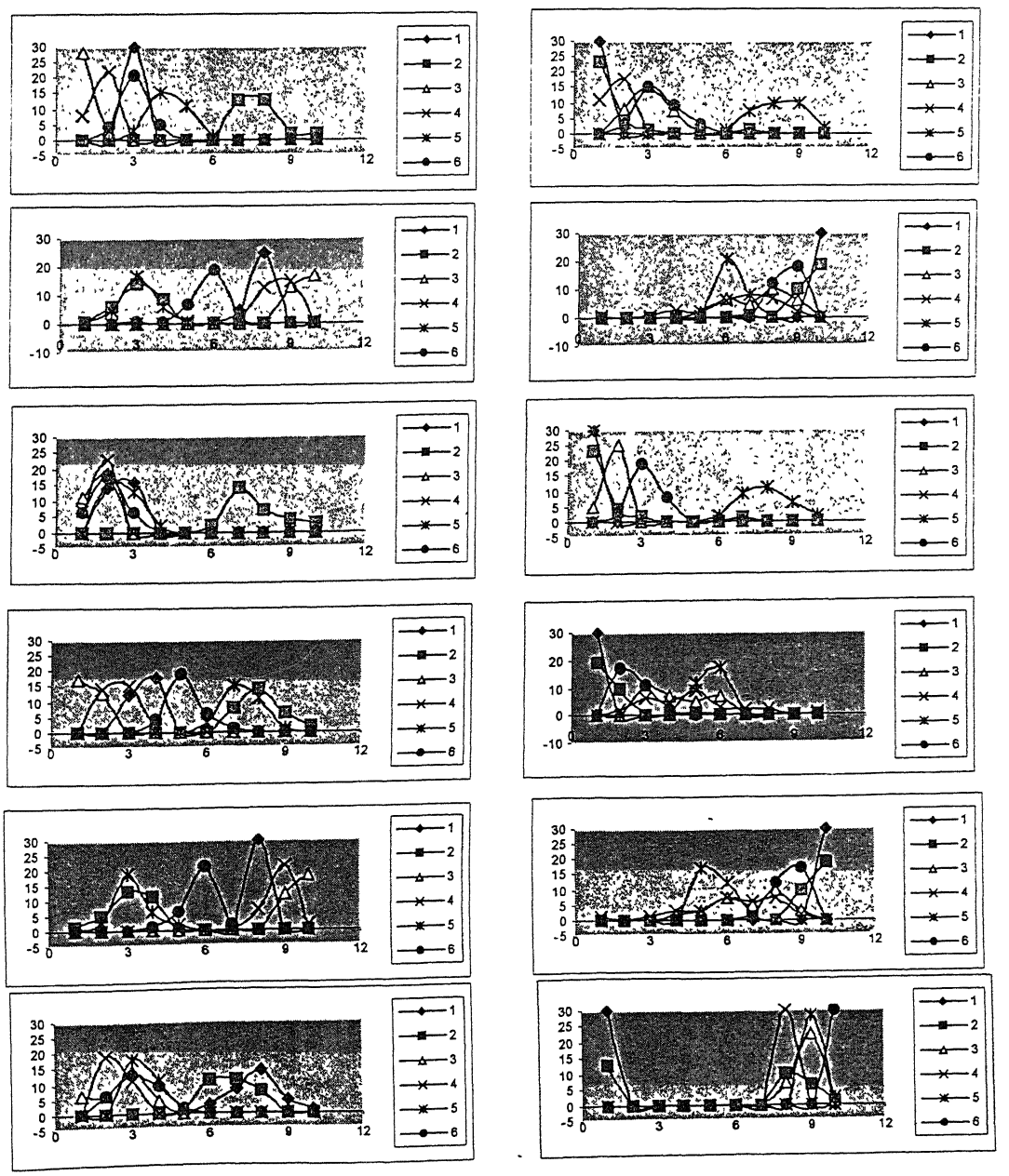
To represent the partial discharge sources in the feature space, in this study, thirty different patterns have been generated for each source. Since the minimum distance classifier and the transformed divergence are based on the assumption that the partial discharge sources are normally distributed, it is very important to investigate the scatter of the partial discharge patterns in the feature space to check the validity of the above assumption. The scatter of the partial discharge sources in the feature space depends upon the features used. The scatter of the six partial discharge sources described in chapter-3 by using all the features of the four texture analysis algorithms are given in this appendix. For the SGLDM, GLDHM, GLRLM and PSM the features are shown in two columns, the first column is for the horizontal and vertical directions of the positive half cycle while the second column is for the features of the horizontal and vertical directions of the negative half cycle. With respect to the q-φ-n distributions, the histograms are shown in three columns, the first one for the features derived from (q_{av}-φ) distribution, the second for the features derived from (q_{max}-φ) distribution while the last one for the features derived from (n-φ) distribution. Generally, in any histograms, the horizontal direction represent the expected values of the sources. While the vertical direction represent the repetition of these values. In this study, each texture analysis technique has features with different scaling. Therefore, for each feature, the minimum and maximum values have been determined to find the range of

variation. This range was then divided into 10 equal intervals and the repetition of the partial discharge patterns of each class in these intervals were determined. Therefore, in the following histograms, the horizontal axis represents the order of the intervals instead of its actual values.

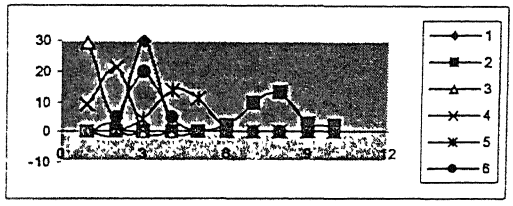
Histograms of the GLDHM features

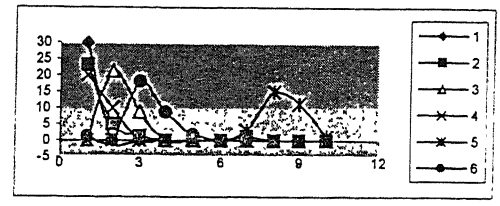
Positive half cycle Negative half cycle

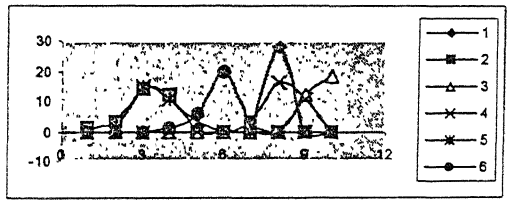
Horizontal direction

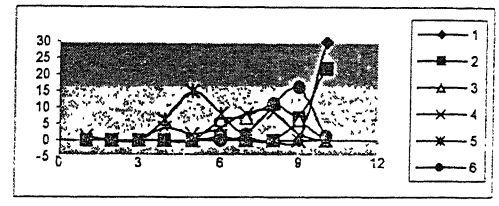


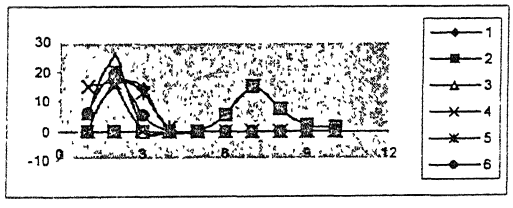
Vertical direction

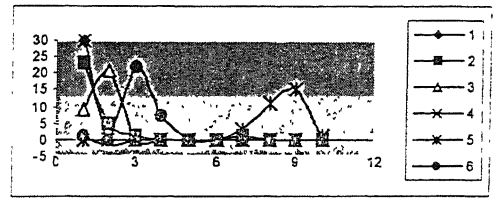


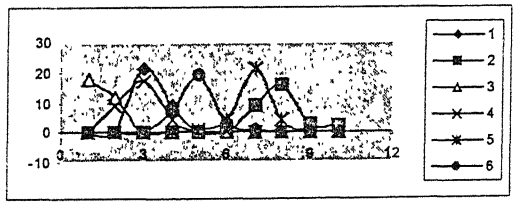


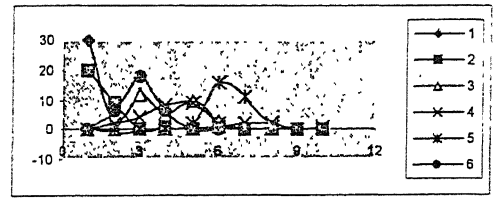


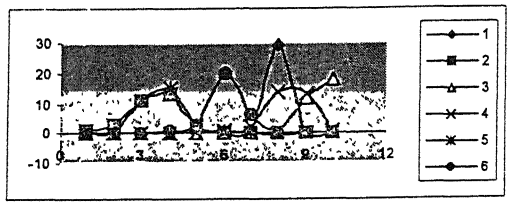


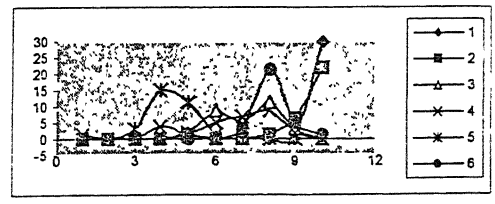










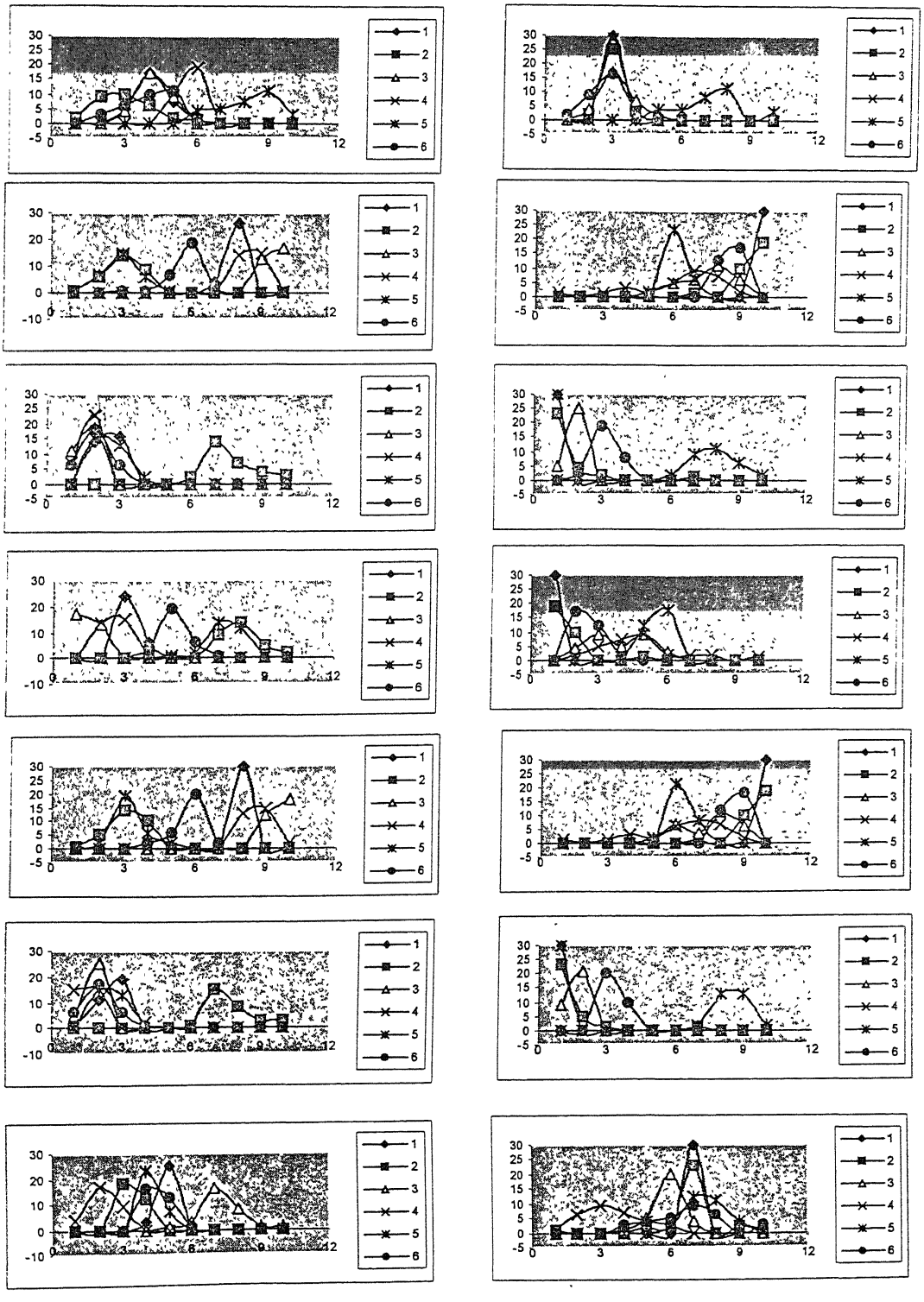


Histograms of the SGLDM features

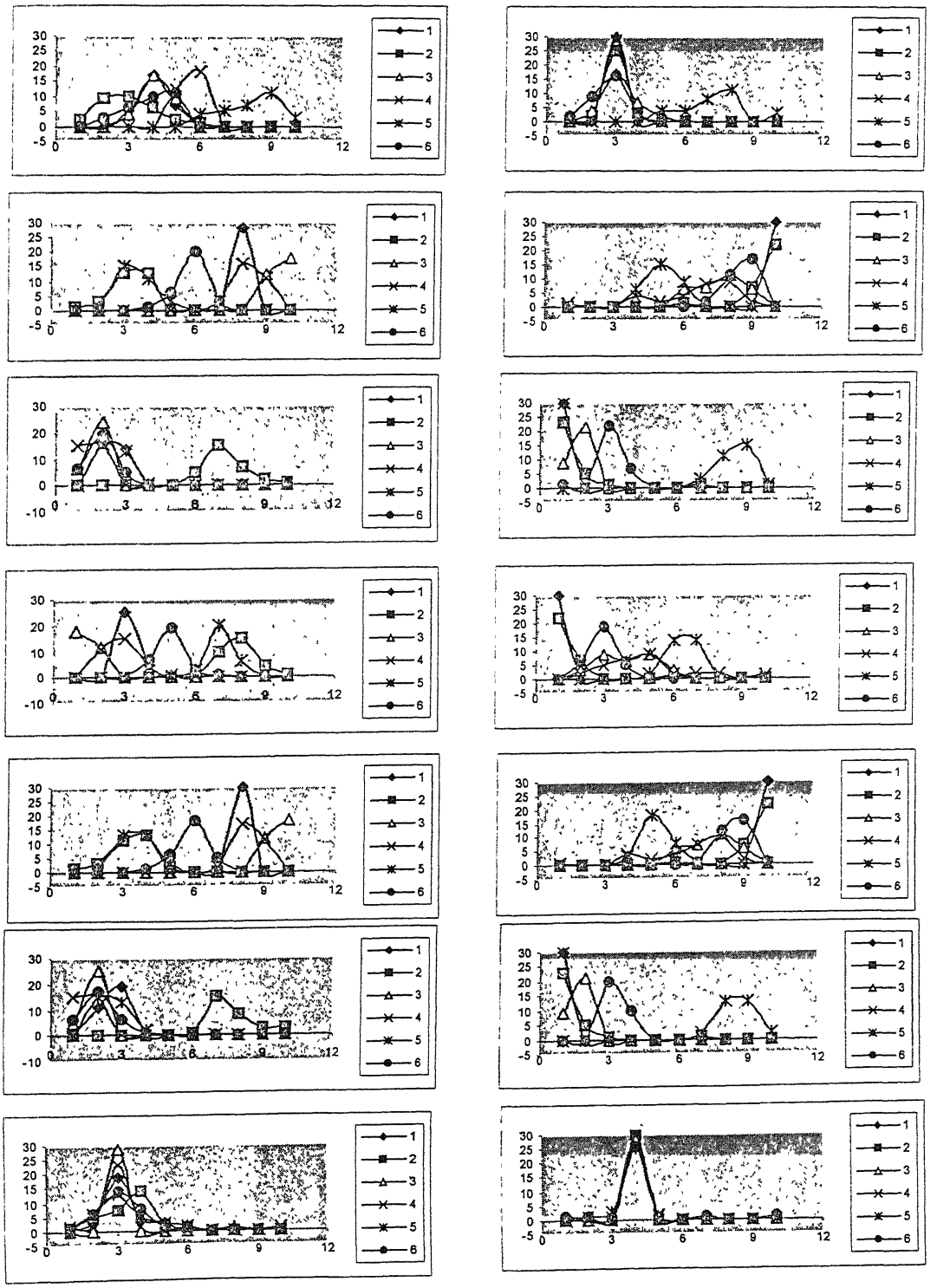
Positive half cycle

Negative half cycle

Horizontal direction



Vertical direction

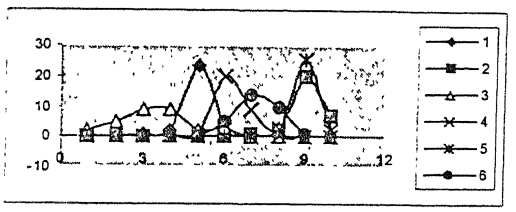


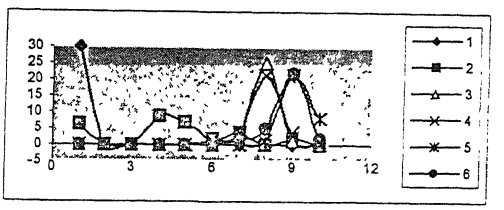
Histograms of the GLRLM features

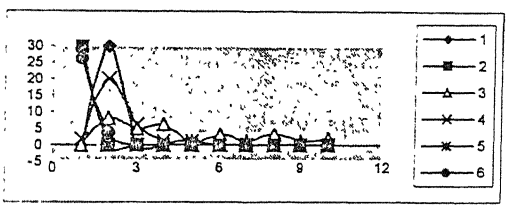
Positive half cycle

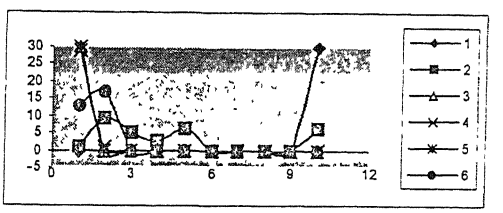
Negative half cycle

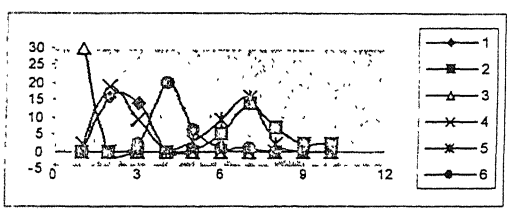
Horizontal direction

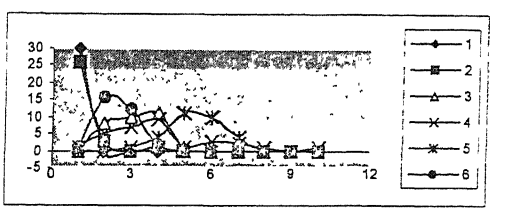


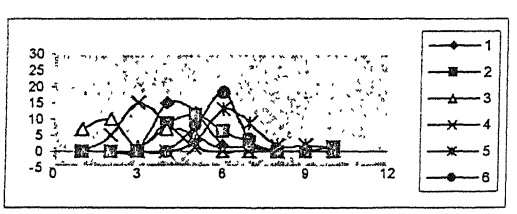


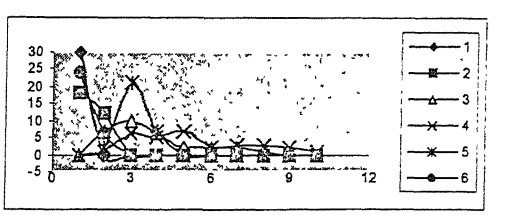


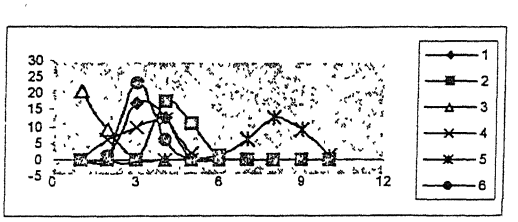


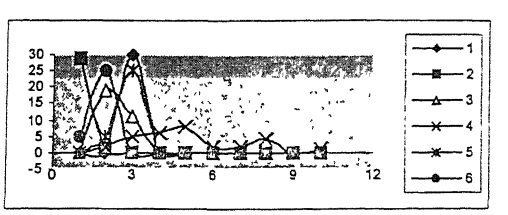




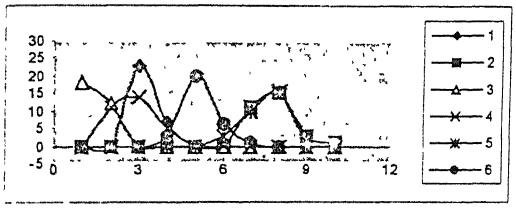


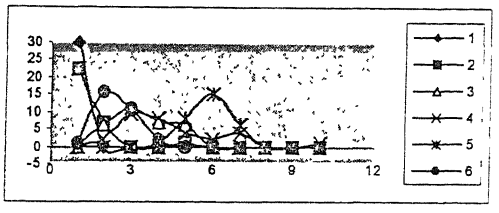


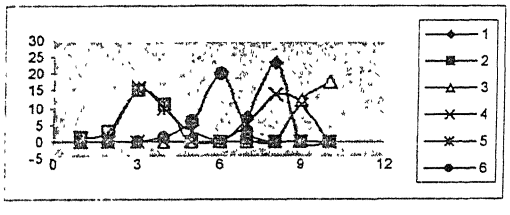


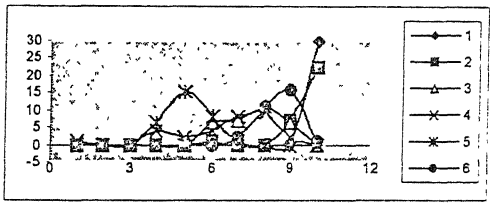


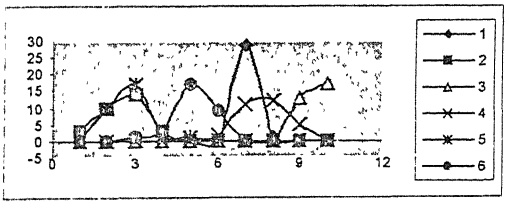
Vertical direction

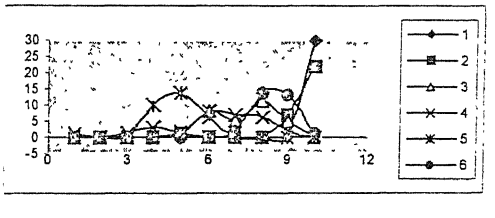


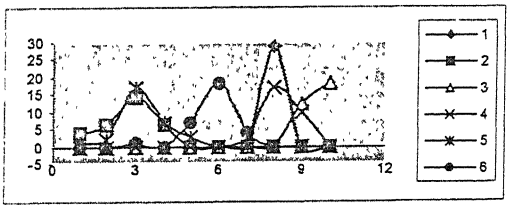


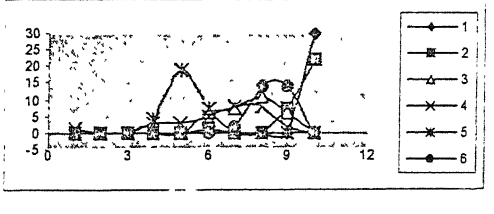


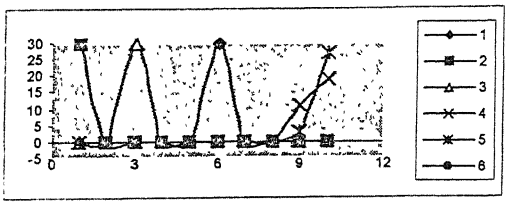


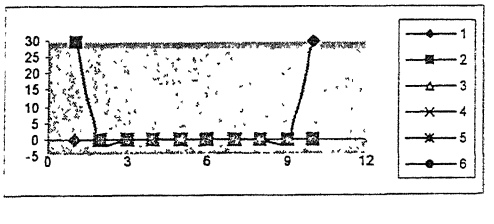










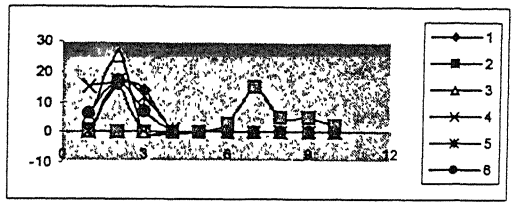


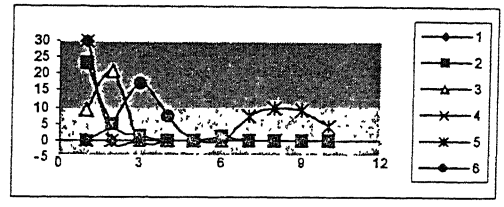
Histograms of the PSM features

Positive half cycle

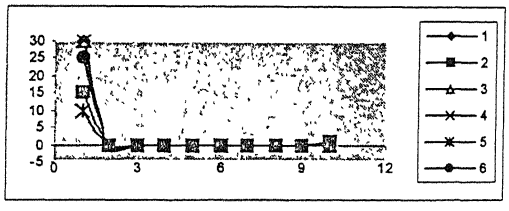
Negative half cycle

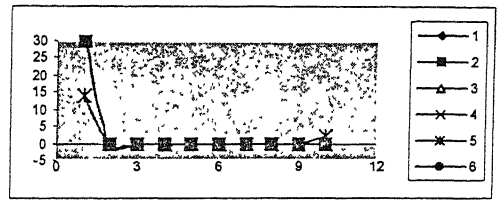
Vertical direction

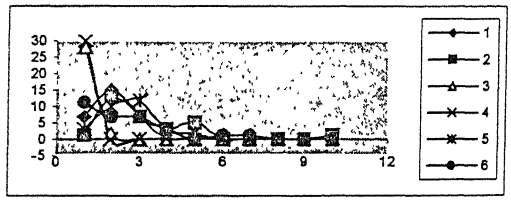


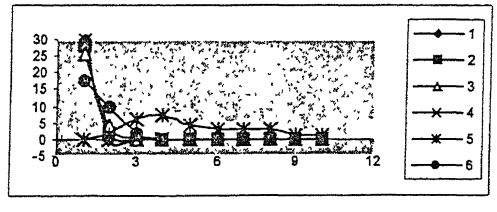


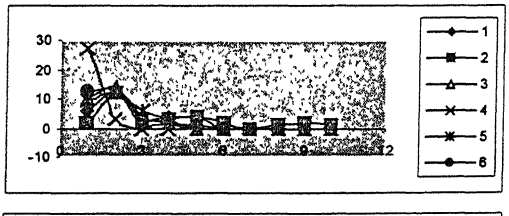
Horizontal direction

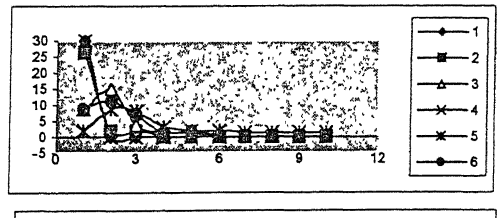


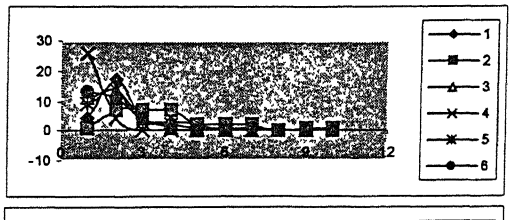


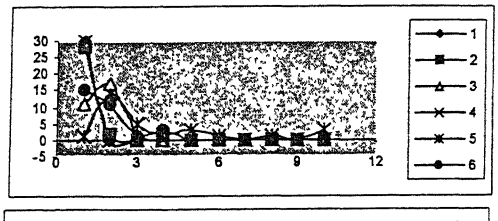


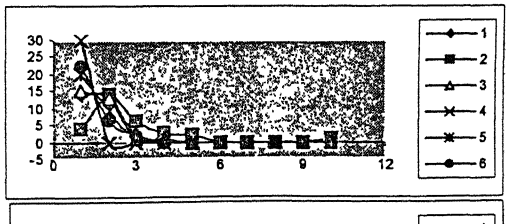


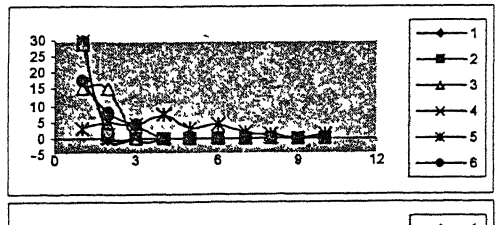


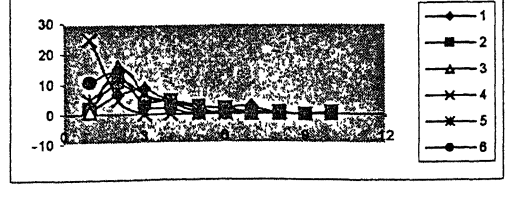


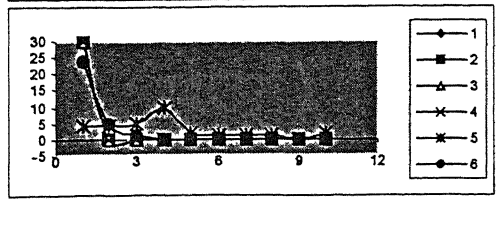






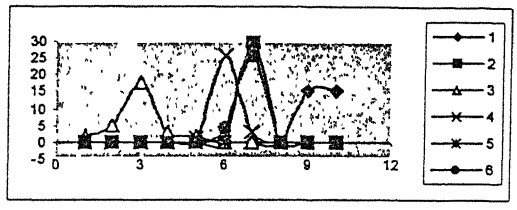


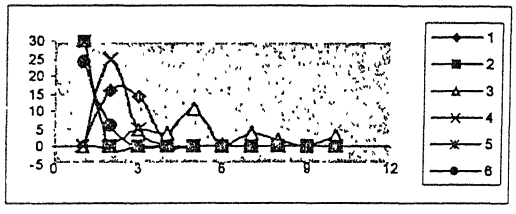


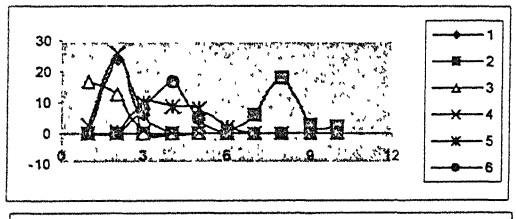


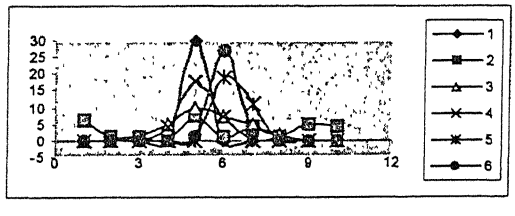
Histograms of the q-φ-n features

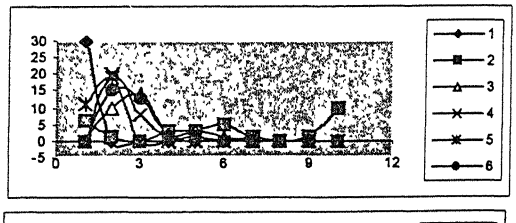
(q_{av} - φ)

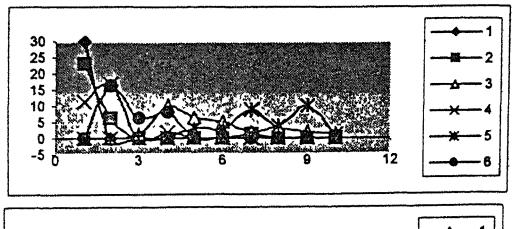


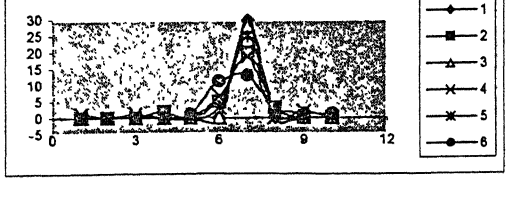




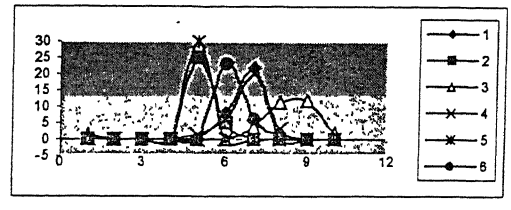


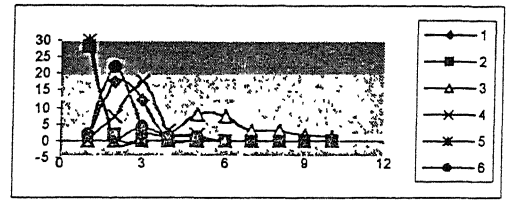


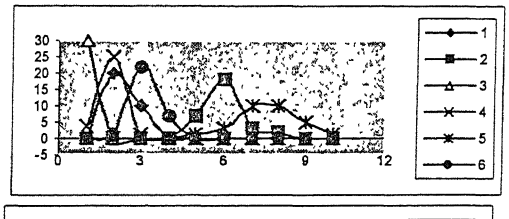


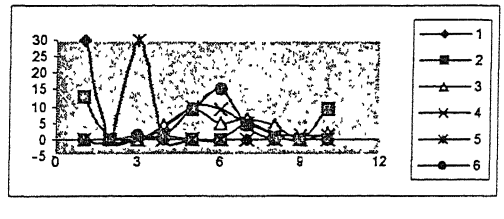


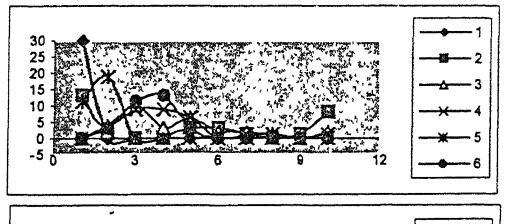
(q_{max} - φ)

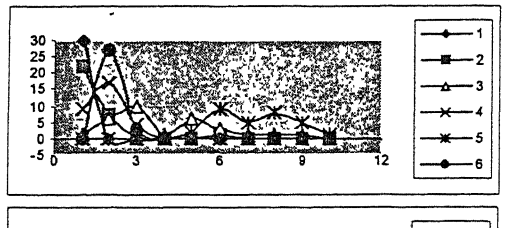


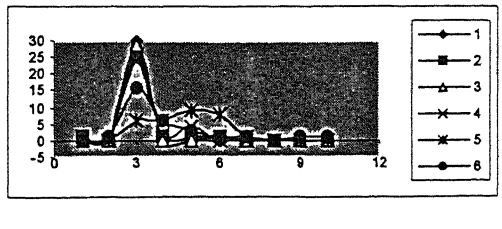




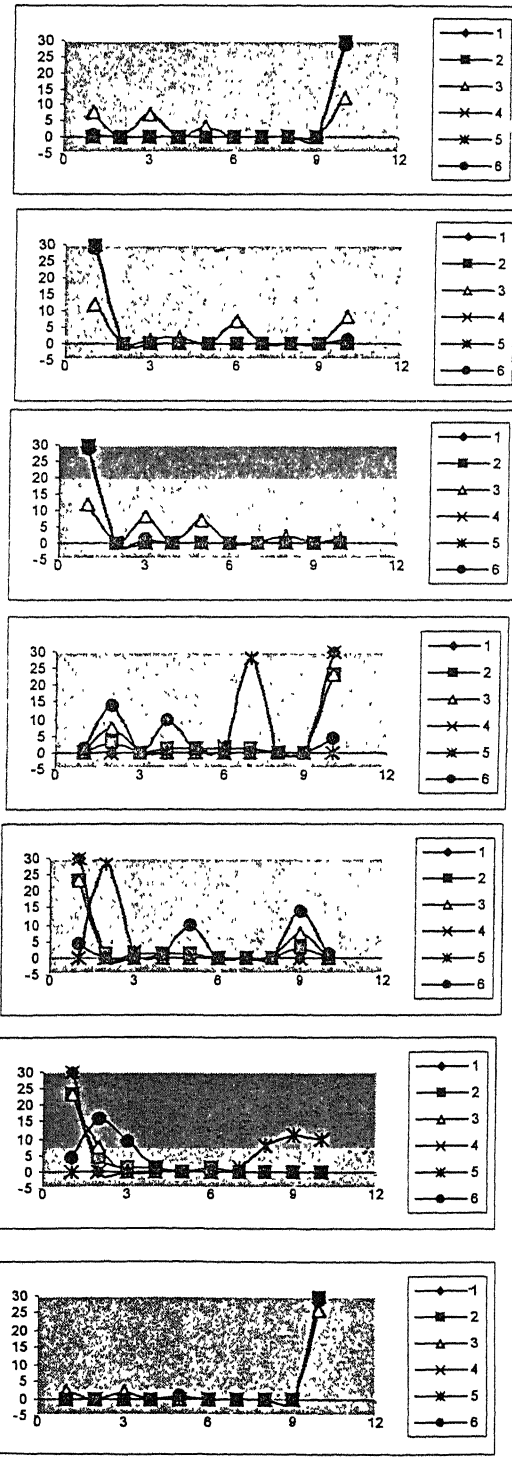








(n - φ)



A131065

Date Slip

This book is to be returned on the date last stamped.